1 GENERAL PROCESSES

2 DIGESTION

3 CIRCULATION

4 BLOOD AND LYMPH

5 RENAL FUNCTION

6 RESPIRATION

7 HUMORAL CONTROL

8 SENSATION

9 NEURAL CONTROL

Best & Taylor's
Physiological Basis of Medical Practice

NINTH EDITION

Best & Taylor's

SECTION EDITORS:

General Physiological Processes
Irving L. Schwartz, M.D.

Digestion
Nicholas C. Hightower, M.D., Ph.D.
Henry D. Janowitz, M.D.

Circulation
David K. Detweiler, V.M.D., D.Sc.

Blood and Lymph
Eugene P. Cronkite, M.D.

The Excretion of Urine
Robert W. Berliner, M.D.

Respiration
William B. Youmans, M.D., Ph.D.
Arthur A. Siebens, M.D.

Endocrine Control Systems
Howard E. Morgan, M.D.

Sensation
John Lott Brown, Ph.D.

Neural Control Systems
John R. Brobeck, Ph.D., M.D.

Physiological Basis of Medical Practice

NINTH EDITION

JOHN R. BROBECK, Ph.D., M.D., Editor

Herbert C. Rorer Professor
in the Medical Sciences, School
of Medicine, University of Pennsylvania,
Philadelphia, Pennsylvania

THE WILLIAMS & WILKINS COMPANY • BALTIMORE, 1973



Copyright © 1973
The Williams & Wilkins Company

All rights reserved This book is protected by copyright No part of this book may be reproduced in any form or by any means including photocopying or utilized by any information storage and retrieval system without written permission of the copyright owner

Made in the United States of America

First Edition—1937
Subsequent Editions—1939, 1943, 1945, 1950, 1955, 1961, 1966
Ninth Edition—1973
Spanish Editions 1939, 1941, 1943, 1947, 1954
Portugese Editions 1940, 1945
Italian Edition 1959
Polish Edition 1960
Rumanian Edition 1966
Indian Edition 1967
Asian Edition 1967

Library of Congress Catalog Card Number 72-85159
SBN 683-10160-3

Reprinted in 1973
Reprinted in 1974

Composed and printed at Waverly Press, Inc
Baltimore, MD 21202 USA

PREFACE

TO FIRST EDITION

Physiology is a science in its own right and the laboratory worker who pursues his researches quite detached from medical problems need offer no apology for his academic outlook Indeed some of the most valuable contributions to medical science have been the outcome of laboratory studies whose applications could not have been foreseen Nevertheless, we feel that the teacher of physiology in a medical school owes it to his students, whose ultimate interest it must be conceded is in the diagnosis and treatment of disease, to emphasize those aspects of the subject which will throw light upon disorders of function The physiologist can in this way play a part in giving the student and practitioner a vantage point from which he may gain a rational view of pathological processes

We have endeavored to write a book which will serve to link the laboratory and the clinic , and which will therefore promote continuity of physiological teaching throughout the pre-clinical and clinical years of the under-graduate course It is hoped that when the principles underlying diseased states are pointed out to the medical student, and he is shown how a knowledge of such principles aids in the interpretation of symptoms or in directing treatment, he will take a keener interest in physiological studies When such studies are restricted to the classical aspects of the subject, apparently remote from clinical application, the student is likely to regard them only as a task which his teachers in their inscrutable wisdom have condemned him to perform Too often he gains the idea, from such a course, that physiology is of very limited utility and comes to believe that, having once passed into the clinical years, most of what he has "crammed" for examination purposes may be forgotten without detriment to his more purely medical studies Unfortunately, he does not always realize at this stage in his education how great has been the part which physiological discoveries have played in the progress of medicine, and that the practice of today has evolved from the "theories" of yesterday

Many physiological problems can be approached only through animal experimentation Advances in many fields, most notably in those of carbohydrate metabolism, nutrition, and endocrinology, bear witness to the fertility of this method of research On the other hand, many problems can be elucidated only by observations upon man, and physiology has gained much from clinical research The normal human subject as an experimental animal possesses unique advantages for many types of investigation, and in disease, nature produces abnormalities of structure and function which the physiological laboratory can imitate only in the crudest way Within recent years the clinical physiologist, fully realizing these advantages and the opportunities afforded by the hospital wards, has contributed very largely to physiological knowledge In many instances, clinical research has not only revealed the true nature of the underlying process in disease, but has cast a light into some dark corner of physiology as well, several examples of clinical investigation which have pointed the way

to the physiologist could be cited In the last century knowledge of the processes of disease was sought mainly in studies of morbid *anatomy*, biochemistry was in its infancy and many of the procedures now commonly employed for the investigation of the human subject had not been devised Today the student of scientific medicine is directing his attention more and more to the study of morbid *physiology* in his efforts to solve clinical problems This newer outlook has borne fruit in many fields It has had the beneficent result of drawing the clinic and the physiological and biochemical laboratories onto common ground from which it has often been possible to launch a joint attack upon disease We feel that this modern trend in the field of research should be reflected in the teaching of medical students, and have therefore given greater prominence to clinical aspects of the subject than is usual in physiological texts

In order to understand the function of an organ it is usually essential to have a knowledge of its structure For this reason we have followed the plan of preceding the account of the physiology of a part by a short description of its morphology and, in many instances, of its nerve and blood supply The architecture and functions of the central nervous system are so intimately related that some space has been devoted to a description of the more important fiber tracts and grey masses of the cerebrum, cerebellum and spinal cord

We wish to thank our colleagues in physiology, biochemistry and anatomy whom we have drawn upon on so many occasions for information and advice without their generous help the undertaking would have been an almost impossible one We are also deeply grateful for the unstinted assistance which we have received from our friends on the clinical staff, several of whom have read parts of the text in manuscript or in proof

C H B
October 15, 1936 N B T

TO NINTH EDITION

We agree In preparing this Ninth Edition we have tried to preserve all of the virtues of earlier editions, and to follow the principles set down by the original authors We have attempted particularly to meet the needs of students in physiology courses in professional schools—medicine, dentistry and veterinary medicine—by offering a volume that is comprehensive but not encyclopedic, and that surveys physiological mechanisms related to problems encountered in clinical practice

Nevertheless, the book has an appearance somewhat different from before The figures are all new and, we trust, more informative Furthermore, the nine sections have been manufactured independently so that they can be revised more easily, and so that they can be published separately as reprints from the complete volume The edge markers serve as guides to section location, and take the place of a continuous pagination

We wish to thank the editors and the staff of The Williams and Wilkins Company for their counsel and cooperation in this effort to present a textbook that is both attractive and easy for readers to use

For the Editors
August 1, 1972 J R B

CONTRIBUTORS

Robert B. Aronson, Ph.D, Associate Chairman, Medical Department, Brookhaven National Laboratory, Upton, Long Island, New York

C. Wayne Bardin, M.D., Associate Professor of Medicine, The Milton S Hershey Medical Center, The Pennsylvania State University, Hershey, Pennsylvania

Robert W. Berliner, M.D., Deputy Director for Science, National Institutes of Health (NIH), Bethesda, Maryland

John R. Brobeck, Ph.D., M.D., Herbert C Rorer Professor in the Medical Sciences, School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania

John Lott Brown, Ph.D., Director, Center for Visual Science, University of Rochester, Rochester, New York

Eugene P. Cronkite, M.D, Chairman Medical Department, Brookhaven National Laboratory, Upton, Long Island, New York, and Professor of Medicine, Health Sciences Center, State University of New York, Stony Brook, New York

David K Detweiler, V.M.D., D.Sc., Professor of Physiology, Head, Laboratories of Physiology, Director, Comparative Cardiovascular Studies Unit, Chairman, Graduate Group in Comparative Medical Sciences, School of Veterinary Medicine, and Graduate School of Arts and Sciences, University of Pennsylvania, Philadelphia, Pennsylvania

Martin Fein, M.S., Fellow, Mount Sinai Graduate School of Biological Sciences of the City University of New York, New York

N. C. Hightower, M.D., Ph.D., Director, Research and Education Division, Scott and White Memorial Hospital, Scott, Sherwood and Brindley Foundation, and the Scott and White Clinic, Temple, Texas

Paula L. Hoffman, M S., Fellow, Mount Sinai Graduate School of Biological Sciences of the City University of New York, New York

Wesley C. Hymer, Ph.D., Associate Professor of Biology, The Pennsylvania State University, University Park, Pennsylvania

Henry D. Janowitz, M.D., Clinical Professor of Medicine, Mt Sinai School of Medicine, New York, New York

Leonard S. Jefferson, Ph.D., Associate Professor of Physiology, The Milton S Hershey Medical Center, The Pennsylvania State University, Hershey, Pennsylvania.

Terry A. Krulwich, Ph D, Assistant Professor of Biochemistry Mount Sinai Medical and Graduate Schools of the City University of New York New York

Sandra K. Masur, Ph.D., Associate, Department of Physiology Mount Sinai Medical and Graduate Schools of the City University of New York New York

Howard E. Morgan, M.D., Professor of Physiology, The Milton S Hershey Medical Center, The Pennsylvania State University Hershey, Pennsylvania

Glenn E. Mortimore, M.D., Associate Professor of Physiology, The Milton S Hershey Medical Center, The Pennsylvania State University, Hershey, Pennsylvania

James R. Neely, Ph.D., Assistant Professor of Physiology, The Milton S Hershey Medical Center, The Pennsylvania State University, Hershey Pennsylvania

John F. Oram, Ph.D., Research Associate in Physiology, The Milton S Hershey Medical Center, The Pennsylvania State University, Hershey Pennsylvania

David N. Orth, M D., Associate Professor of Medicine Vanderbilt University, Nashville, Tennessee

Zia J. Penefsky, Ph.D., Assistant Professor, Department of Physiology Mount Sinai Medical and Graduate Schools of the City University of New York, New York

Irving L. Schwartz, M.D., Professor and Chairman, Department of Physiology, Mount Sinai School of Medicine, Dean, Mount Sinai Graduate School of Biological Sciences of the City University of New York, Research Collaborator, Brookhaven National Laboratory, Upton, Long Island, New York

Arthur A. Siebens, M.D., Physician-in-Chief, Department of Rehabilitation Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland

George J. Siegel, M.D., Associate Professor, Departments of Neurology and Physiology, Mount Sinai Medical and Graduate Schools of the City University of New York, New York

William E. Stone, Ph.D., Professor of Physiology, University of Wisconsin Medical School, Madison, Wisconsin

Ralph A. Straffon, M.D., Head, Department of Urology, Cleveland Clinic, Cleveland, Ohio

Anthony C. Trakatellis, M.D., Ph.D., Professor of Biochemistry, Mount Sinai Medical and Graduate Schools of the City University of New York, New York

C. Robert Valeri, M.D., Officer in Charge, Naval Blood Research Laboratory, U S Naval Hospital, Chelsea, Massachusetts

Roderich Walter, Ph.D., Professor of Physiology, Mount Sinai Medical and Graduate Schools of the City University of New York, Research Collaborator, Brookhaven National Laboratory, Upton, Long Island, New York

Carol F. Whitfield, Ph.D., Assistant Professor of Physiology, The Milton S. Hershey Medical Center, The Pennsylvania State University, Hershey, Pennsylvania.

Herman R. Wyssbrod, Ph.D., Assistant Professor, Departments of Physiology and Biophysical Chemistry, Mount Sinai Medical and Graduate Schools of the City University of New York, New York.

William B. Youmans, M.D., Ph.D., Professor of Physiology, University of Wisconsin Medical School, Madison, Wisconsin.

CONTENTS

Section 1
GENERAL PHYSIOLOGICAL PROCESSES
Edited by Irving L Schwartz

1 **By Way of Introduction—The Cell** 1-1
By Sandra K Masur, Roderich Walter and Irving L Schwartz

2 **Physicochemical Properties, Permeability, and Transport Functions of Cell Membranes** 1-15
By Irving L Schwartz Roderich Walter Martin Fein and Herman R Wyssbrod

3 **Excitation, Conduction, and Transmission of the Nerve Impulse** 1-29
By Irving L Schwartz and George J Siegel

4 **Excitation and Contraction of Muscle** 1-61
By Irving L Schwartz and Zia J Penefsky

5 **Amino Acids and Proteins The Molecular Framework and Machinery of Living Systems** 1-114
By Irving L Schwartz and Martin Fein

6. **Physiology of Purines and Pyrimidines** 1-142
By Roderich Walter and Paula L Hoffman

7 **Nucleic Acids and Gene Activity** 1-158
By Anthony C Traketellis (with the assistance of) Terry Krulwich

Section 2
DIGESTION
Edited by N C Hightower, Jr , and Henry D Janowitz

1. **Control of Food and Water Intake** 2-1

2 **Innervation and Visceral Sensations of the Gastrointestinal Tract** 2-8

3. **Gastrointestinal Hormones** 2-15

4 **Salivary Secretion** 2-17

5. **Gastric Secretion** 2-31

6. **Pancreatic Secretion** 2-49

7 **Biliary Secretion** 2-60

8 **Secretion and Absorption in the Intestine** 2-79

9 **Movements of the Alimentary Canal** 2-95

Section 3
CIRCULATION
Edited by David K Detweiler

1	The Cardiovascular System	3 1
2	The Heart Gross Structure, Myocardial Cells	3 25
3	Electrophysiology of the Heart	3 48
4	Mechanical Activity of the Heart	3 86
5	Regulation of the Heart	3 110
6	The Coronary Circulation	3 130
7	Measurement of Blood Pressure and Flow	3 148
8	Control Mechanisms of the Circulatory System	3 164
9	Regulation of Systemic and Pulmonary Circulation	3 189
10	Circulation through Brain, Skin, and Skeletal Muscle	3 211
11	Splanchnic, Renal, and Fetal Circulations	3 224

Section 4
BLOOD AND LYMPH
Edited by Eugene P Cronkite

1	Blood and Lymph General Properties, Composition and Functions	4-1
2	Hemopoiesis A Vital Cell Renewal System	4 10
3	The Erythrocyte	4 21
4	Hemoglobin By C Robert Valeri	4 41
5	Transfusions The Blood Groups By C Robert Valeri	4 48
6	Hemolysis and Suspension Stability of the Blood By C Robert Valeri	4 57
7	Granulocytes	4 62
8	Thrombopoiesis, Hemostasis, Coagulation, and Thrombosis	4 73
9	The Lymphatics, Lymph, and Tissue Fluids	4-83
10	Lymphoreticular System and the Lymphocyte	4-90
11	Regulation of the Hydrogen Ion Concentration of Body Fluids By Robert A Aronson	4-103
12	Fluid Distribution and Exchange	4-113

Section 5
THE EXCRETION OF URINE
Edited by Robert W Berliner

1	Urine Formation	5-1
2	Pathological Physiology of the Kidney. Micturition By Ralph A Straffon	5-38

Section 6
RESPIRATION
Edited by William B Youmans and Arthur A Siebens

1 **Uptake and Delivery of the Respiratory Gases** 6-1
By William E Stone

2 **Pulmonary Gas Exchange** 6-20
By Arthur A Siebens

3 **The Mechanics of Breathing** 6-28
By Arthur A Siebens

4 **Control of Breathing** 6-41
By William B Youmans

5 **Hypoxia, Asphyxia, Dysbarism, Oxygen Therapy, and Resuscitation** 6-63
By William B Youmans

Section 7
ENDOCRINE CONTROL SYSTEMS
Edited by Howard E Morgan

1 **Introduction to Endocrine Control Systems** 7-1
Howard E Morgan

2 **Control of the Anterior Lobe of the Pituitary Gland** 7-8
Wesley C Hymer

3 **Control of the Posterior Lobe of the Pituitary Gland** 7-24
Howard E Morgan

4 **Control of Thyroid Function** 7-29
Glenn E Mortimore

5 **Control of Calcium Homeostasis by Parathyroid Hormone, Calcitonin, and Vitamin D** 7-42
Carol F Whitfield

6 **Function of the Adrenal Glands** 7-50
David N Orth

7. **Hormonal Control of the Pancreatic Islets** 7-76
Glenn E Mortimore

8. **Sexual Differentiation** 7-88
C Wayne Bardin

9 **Hormonal Control of Gonadal Function** 7-95
C Wayne Bardin

10 **Control of Carbohydrate Metabolism in Muscle** 7-112
James R Neely

11 **Control of Fatty Acid Metabolism in Adipose Tissue** 7-123
James R Neely and John F Oram

12 **Control of Gluconeogenesis, Ketogenesis, and Lipogenesis in Liver** 7-132
Leonard S Jefferson

13 **Hormonal Control of Growth and Protein Metabolism** 7-143
Howard E Morgan

Section 8
SENSATION
Edited by John Lott Brown

1	Sensory Systems	8 1
2	Hearing	8 10
3	Vestibular Function	8 40
4	Vision	8 63
5	Taste	8 121
6	Olfaction	8 132
7	Touch, Temperature, and Pain Sensitivity	8 144

Section 9
NEURAL CONTROL SYSTEMS
Edited by John R Brobeck

1	Analysis of Neural Control Systems	9 1
2	Organization of the Central Nervous System	9 20
3	Control of Smooth Muscle and Exocrine Glands	9 53
4	Control of Skeletal Muscle	9 66
5	Subcortical Mechanisms Controlling Posture and Locomotion	9 88
6	Cortical Control Systems	9 104
7	Control Systems That Establish Regulations	9 121
8	Higher Neural Functions	9 137
9	Disorders of Neural Control	9 161
	INDEX	

1

General Physiological Processes

1 / BY WAY OF INTRODUCTION—THE CELL

A multicellular organism is a society of cells which like the units of any society, manifest division of labor, specialization, and mutual interaction according to a set of rules and centralized controls Consider the levels of organization of the human organism *Atoms* are combined in different ways to yield *monomers* (amino acids, nucleotides, sugars) which are the building blocks of biological *polymers* (proteins, nucleic acids, polysaccharides) These macromolecules are combined with smaller molecules to form *organelles* (subcellular units with distinctive function and morphology), which in turn are combined to form cells The study of mammalian physiology involves the dynamic interrelationships among cells, tissues, and organs and reaches ultimately to the level of the organism as a whole In this chapter we will consider the structure and function of the organelles which comprise the cell One group of organelles is bounded by a limiting membrane, a second group is not so delimited The former group encompasses the nucleus, endoplasmic reticulum, Golgi apparatus, lysosome, peroxisome, and mitochondrion—and in this category we also include the plasma (outer) membrane of the cell as a whole The group of organelles that are not bounded by a membrane includes the chromosomes, nucleoli, microtubules, microfilaments, and centrioles

Although there was some knowledge of the cell boundary and subcellular organelles prior to the advent of the electron microscope, the complicated ultrastructure of intracellular membrane systems was unforeseen (Fig 1 1)

The following two membrane properties are particularly significant in accounting for differences of function among the various membrane-delimited subcellular organelles *1*) the property of selective permeability which establishes and maintains compartments of specific character within the cell, and 2) the ability to provide a matrix which lends its surface as well as its interior for the arrangement of an ordered sequence of molecules (enzymes, cofactors, carriers) in a functionally meaningful pattern

Cell Membrane

The cell membrane is a permeability barrier If a cell is placed in hypotonic solution and if it contains molecules which cannot penetrate its outer membrane, it will swell, conversely it will shrink if placed in a hypertonic medium In both instances water moves down its concentration gradient (chap 2) Thus, the cell behaves as an osmometer Nonpolar molecules (gases, lipids) move freely across the membrane, polar molecules penetrate the membrane much less readily and indeed it is the selective permeability of the plasma membrane to certain ions which determine the excitability characteristics of nerve and muscle cells

COMPOSITION OF THE CELL MEMBRANE

Although the general chemical nature of the membrane found at the cell boundary was predicted on the basis of physiological data, the detailed molecular structure of the membrane is not yet known Models of cell membranes prepared by combining their lipid and protein constituents (partly known from chemical analysis of purified cell membrane preparations) exhibit some physiological characteristics similar to those of natural membranes The role of lipid-lipid interactions in membrane structure has been at the center of attention because such interactions can explain much of the presently known phenomena of membrane transport Quantitative studies of isolated cell membranes revealed that enough lipid is present to be arranged as a bilayer coating the cell Artificial mixtures of extracted cellular polar lipids (lecithin, phospholipid, and steroids) under appropriate conditions will form a biomolecular layer

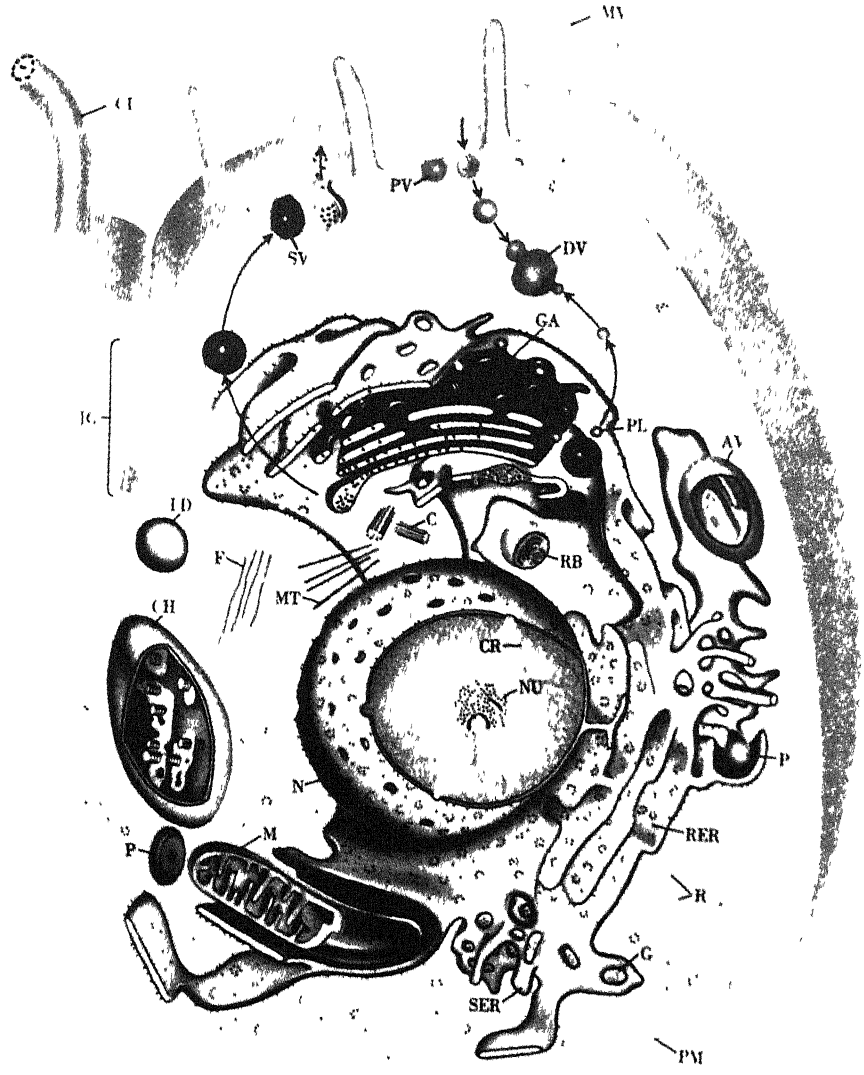


Fig 1 1 Schematic diagram of a cell and its organelles drawn to reveal their three dimensional structure *AV*, autophagic vacuole *C* centriole *CH* chloroplast *CI*, cilium *CR* chromatin *DV* digestion vacuole *F* microfilaments *G* glycogen *GA* Golgi apparatus, *JC* junctional complex *LD* lipid droplet *M* mitochondrion *MT* microtubules *MV* microvillus *N* nucleus *NU* nucleolus *P* peroxisome *PL* primary lysosome, *PM*, plasma membrane *PV* pinocytosis vesicle *R* ribosomes and polysomes *RB* residual body, *RER* rough endoplasmic reticulum *SER* smooth endoplasmic reticulum *SV* secretion vacuole The organelles have been drawn only roughly to scale Also the sizes and relative amounts of different organelles can vary considerably from one cell type to another For example only plant cells show chloroplasts and among animal cells only a few types show peroxisomes (× 58 000) (From Novikoff and Holtzman 1970)

spontaneously Presumably the polar (hydrophilic) ends of the lipids form the two outer borders making them available for interaction with other polar molecules such as proteins (Fig 1 2) On a weight basis membranes contain a significantly larger amount of protein than lipid (ratio up to 4 1), however, due to the high molecular weight of proteins, this relationship is reversed on a molar basis (protein· lipid ranging from 1 to 10·100) At least some of the protein must be present in a globular rather than extended form, unfolding of enzymic protein results in loss of activity In fact, some of the proteins are probably not restricted to the surfaces of the plasma membranes but rather extend into the bimolecular lipid layer

In addition to lipid and protein, carbohydrate appears to be associated with the cell membrane as lipopolysaccharide and as protein-polysaccharide The carbohydrate moieties of the membrane serve to modify the electrical charge at its surface and provide specific surface-binding sites Cytochemically demonstrable polysaccharide and protein is associated with many cell surfaces as an extracellular layer In some places, particularly at luminal surfaces this layer forms a fuzzy coat—often referred to as a glycocalyx (Fig 1 3)—which may act as a crude filter and/or facilitate the attachment of molecules for endocytic (see below) transport across the cell membrane

STRUCTURE OF UNIT MEMBRANE

A pattern generally found in almost all cellular membranes prepared for microscopy by conventional techniques consists of three layers, two electron-dense layers on either side of a single electron-lucent layer (Fig 1 3 and 1 4) This is termed a *unit membrane* Electron microscopic examination of osmium tetroxide-fixed, sectioned tissue shows the cell membrane to be 70 to 100 Å wide The electron-lucent line in the unit membrane is thought to represent the lipid layer Hydrophobic bonding in the lipid bilayer region may make it inaccessible to osmium deposition Thus, the two electron-dense lines would result from deposition of osmium at the surfaces of this bilayer There is much physiological evidence to suggest that the lipid bilayer is interrupted by hydrophilic molecules (Fig 1 2*A*) which connect the two outer surfaces of the membrane and provide transmembrane channels (a few Å in width) for transfers of water molecules and ions (chap 2)

Evidence for the unit membrane structure has been provided by x-ray diffraction patterns

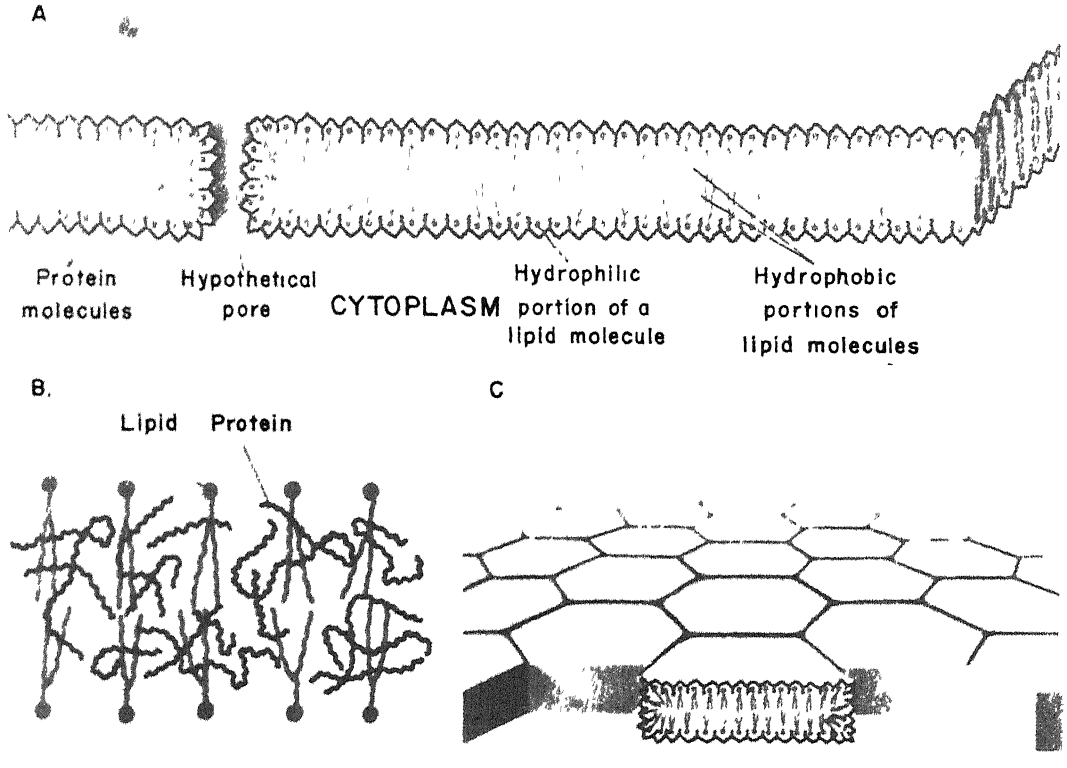


Fig 1.2 Diagram of various models of the cell membrane *A* represents the classic biomolecular lipid leaflet coated by protein note inclusion of a hypothetical hydrophilic pore *B* represents a less ordered protein component found within the lipid bilayer *C* represents a mosaic of subunits each made up of lipid bilayers with protein or other hydrophilic molecules forming the borders All of these forms of the membrane (and others) could be present simultaneously and in different proportions under different conditions (From Novikoff and Holtzman, 1970)

of the myelin sheath This membrane-derived structure lends itself well to x-ray studies because of its jelly roll nature (Fig 1 4) which gives rise to a highly ordered and repetitive form Indeed, the myelin diffraction pattern can be explained on the basis of repeating bilayers of lipids in the cell membrane

Additional structural information comes from unfixed membranes studied by the freeze-etch technique This technique involves the quick freezing and sublimation of unfixed membranes Membranes so treated tend to fracture along the middle layer of the unit membrane as would be expected from a bimolecular lipid leaflet Moreover, repeating structures are visible at various intervals within the layer, these structures may represent proteins that extend through much of the thickness of the membrane

STRUCTURAL ASPECTS OF TRANSPORT ACROSS THE MEMBRANE

Although the physiology of the movement of molecules across membranes is dealt with in Chapter 2, we will here indicate a few selected examples of transport phenomena correlated with structural features of the cell membrane The lipid bilayer serves as a transport medium for lipid-soluble molecules to gain entry into the cell, whereas, protein-lined hydrophilic "pores" probably provide channels for diffusion of polar entities such as water and ions Specialized membrane-bound proteins seem to be involved in enzymatically controlled transport

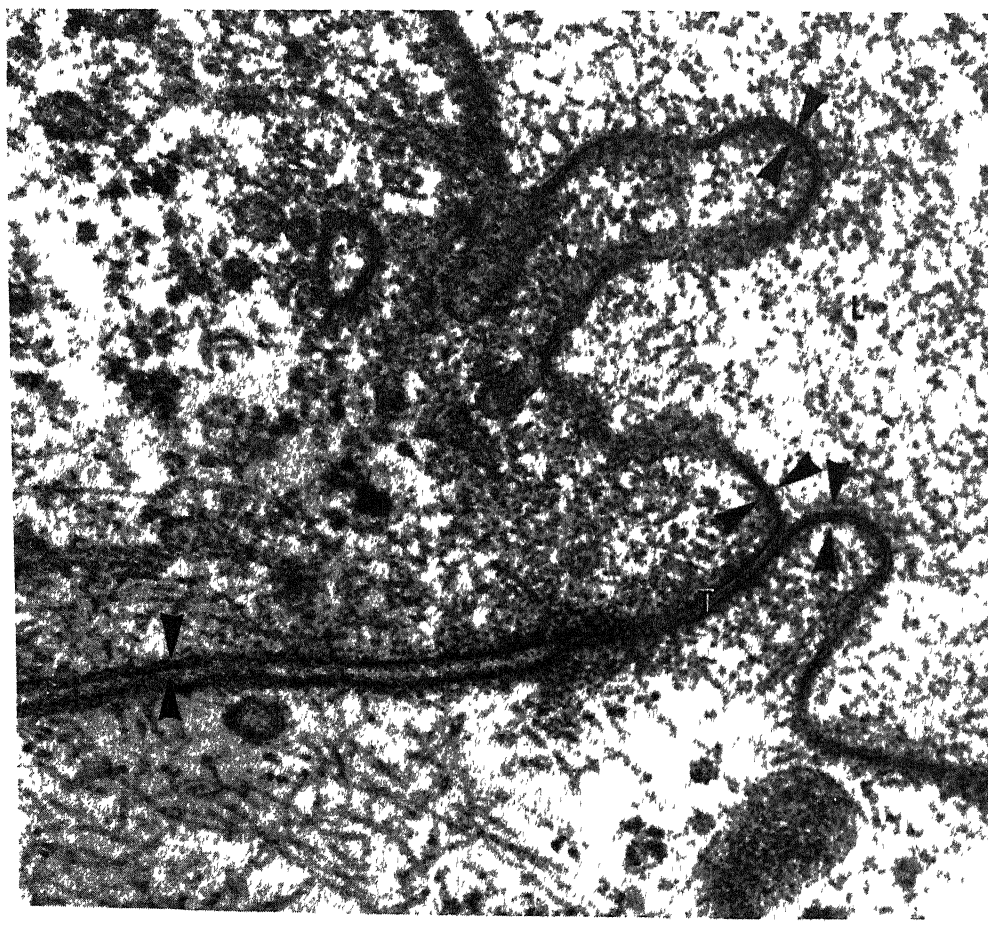


Fig 1 3 Electron micrograph of portions of epithelial cells (E) bordering the lumen (L) of the toad urinary bladder The unit membrane structure of two cell membranes is seen between the arrows In the tight junction (T) between the cells the external surfaces of the two cell membranes appear to have fused so that three rather than four dense lines are seen The fuzz in the lumen is associated with the cell surface (see text) (× 58,000)

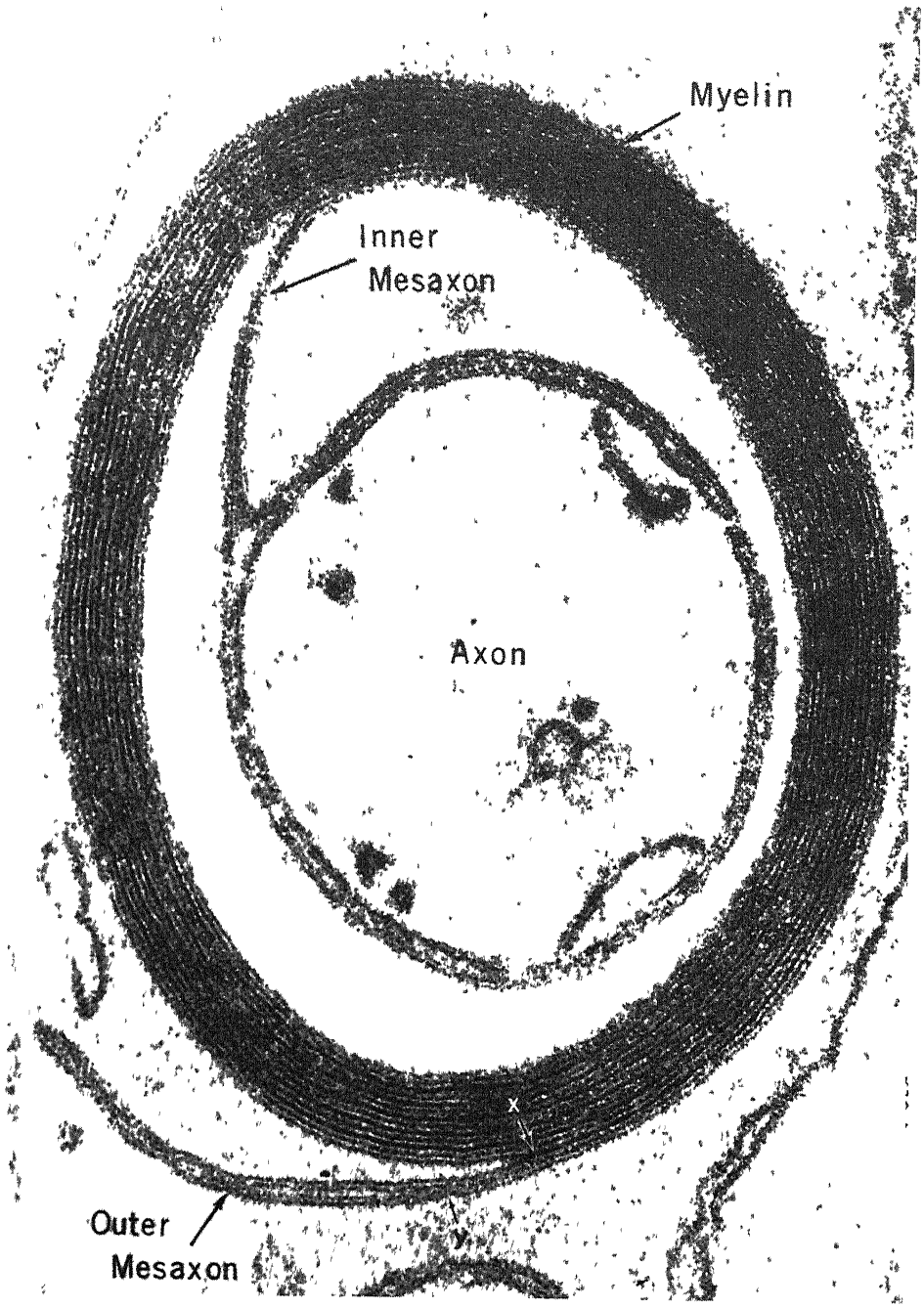


Fig 1 4 Electron micrograph of cross section of axon surrounded by the Schwann cell including the myelin sheath The unit membrane of the axon and the inner and outer mesaxon can be seen surrounding the axon At the arrow marked *x*, the cytoplasmic surfaces of the two Schwann cell membranes come together to form the denser line and at the arrow marked *y*, the outer surfaces of the two membranes of the outer mesaxon come together to form the thinner line (Potassium permanganate fixation, × 193 000) (Robertson from Fawcett 1966)

processes. For example a particular type of adenosine triphosphatase (ATPase) is involved in the active transport of Na^+ and K^+ ions, a variety of "permeases" are involved in the initiation of transmembrane sugar transport, and adenylcyclase, a membrane-bound enzyme that

converts adenosine triphosphate (ATP) to 3′-5′-cyclic adenosine monophosphate (AMP), serves to mediate the actions of several hormones which attach to "receptor sites" on the surface of their target cells

STRUCTURAL ASPECTS OF TRANSPORT OF THE MEMBRANE

Membrane components are subject to continual turnover In certain cell types, portions of the membrane invaginate into the cell and pinch off to form the boundary of an intracellular vesicle, vacuole, or tubule External material is carried into the cell by this process, referred to as *endocytosis* (Fig 1 1, see also chap 2) This material and its enclosing membrane generally fuses with lysosomes (see below) *Exocytosis*, on the other hand, involves the fusion of membranes of intracellular origin, such as those delimiting secretory granules, with the cell membrane These granules release their internal material to the outside of the cell As a result there is a flow of membrane and of material enclosed in membrane-delimited spaces between the surface and intracellular compartments Similar exchanges seem to occur between certain intracellular organelles, notably the endoplasmic reticulum and the Golgi apparatus

STRUCTURAL ASPECTS OF INCREASING THE CELL SURFACE

Stable evaginations and invaginations of the cell membrane are important elements in providing a dramatic increase in surface area contact between cell and environment Microvilli, finger-like evaginations, are generally associated with cell surfaces involved in absorption processes such as in the intestine and kidney Conversely, in striated muscle, one finds an invagination of the cell membrane, the transverse tubule, associated with each sarcomere Since the transverse tubules are continuous with the surface membrane they provide a direct route for the communication of alterations at the cell surface to the contractile system deep within the muscle fiber (chap 4)

The Nucleus

FUNCTION OF NUCLEUS

The nucleus has two principal functions replication of deoxyribonucleic acid (DNA) and synthesis of ribosomal messenger and transfer ribonucleic acid (RNAs) Because it is best understood, we shall discuss in some detail ribosomal RNA synthesis which occurs in the nucleolus

Each nucleus possesses one or more *nucleoli* not delimited by membranes Each nucleolus consists of a roughly spherical, dense array of fibrils and granules rich in RNA Often the nucleolus is found in intimate association with special regions of DNA (known as nucleolar organizer regions) which are presumed to carry the information for *ribosomal RNA* (rRNA) Nucleolar RNA (45S predominantly) is almost certainly a precursor form of ribosomal RNA found in the cytoplasm if one "labels" RNA synthesized in the nucleolus with radioactive nucleotides, labeled RNA molecules are subsequently detected in the cytoplasm These RNA molecules (18S and 28S) complex with protein and form, respectively, a 40S and a 60S subunit One small and one large subunit combine to form a *ribosome* (80S) 150 to 250Å in diameter There may be several million ribosomes in a given cell

The specific function of ribosomal RNA is not well-understood, generally speaking however, all three types of RNA are involved in the translation of genetic information contained in the DNA molecule into specific proteins, synthesized in the cytoplasm (chap 7) The ribosomes interact in the process of protein synthesis with two other types of RNA large *messenger RNA* molecules (mRNAs) determine the sequence of amino acids in proteins by specifying the order of attachment of the small *transfer-RNAs* carrying the appropriate amino acids Our belief that DNA is the template for these RNAs is derived largely from experiments with procaryotic (bacterial) cells

As the information in a messenger RNA molecule (mRNA) is being "read," several ribosomes attach via their smaller subunits to the mRNA The combination of an mRNA and its attached ribosomes is referred to as a *polysome*. The nascent peptide seems to be attached to the larger ribosomal subunit (60S), completed protein is released to the cytoplasm (chap 7)

MORPHOLOGY OF THE NUCLEUS

The interphase nucleus is readily seen in the light microscope as a spheroidal body with a "suggestion" of internal organization (Fig. 1 1)

The DNA containing material can be specifically stained The nuclear chromatin can be resolved into two types euchromatin (loosely coiled) and heterochromatin (compact) It seems likely that the euchromatic regions are more active in the transcription process than the heterochromatic regions, i e , there is little demonstrable RNA synthesis in chromosomes that are largely or entirely composed of heterochromatin as in the case of sperm cells, polymorphonuclear leukocytes and the Barr body (one of the X chromosomes of female cells) The association of euchromatin with active transcription may account for some of the selective genetic expression associated with characteristic chromosomal uncoiling patterns found in different tissues within the same organism or at different developmental stages in the same tissue

Chemical analysis has shown that the chromosome consists of DNA associated with basic proteins (histones) and with other (nonhistone) protein It has been speculated that the complexing of histone with DNA may have a protective or structural function (preventing alteration or denaturation of the DNA, controlling coiling, etc) or the histone may have a repressor function (interfering with the template activity of DNA)

Isolated chromosomes studied by electron microscopy appear as masses of fibers around 250 Å in diameter, but the architecture of these fibers is not known An individual DNA double helix coated with protein measures less than 50 Å and while it is known that the fibers of chromosomes are coiled, the nature of the packaging of nucleic acid and proteins is yet to be described. Nevertheless, there are theories consistent with current evidence, suggesting that a single chromosome contains one, or at most a very few, extremely elongated DNA molecules complexed with protein and coiled into a fiber structure which is seen in the electron microscope.

Intracellular Membrane Systems

THE NUCLEAR ENVELOPE

The boundary of the nucleus, the nuclear envelope, is a double membrane complex (Fig 1 5) Each membrane is approximately 70 to 80 Å thick The envelope, a flattened sac with an enclosed perinuclear space, resembles the rough endoplasmic reticulum (ER) (see below) *1*) the cytoplasmic surface of the outer (cytoplasmic) nuclear membrane has granules which appear to be ribosomes, *2*) direct continuities are seen between the cytoplasmic portion of the nuclear membrane and the ER (Fig 1 5), and *3*) the presence of certain enzymes can be demonstrated cytochemically in both the perinuclear space and the cisternae of the ER

The inner surface of the nuclear membrane is often associated with chromatin and an "internal dense lamella," the latter may provide some rigidity to the structure The inner and outer membranes of the envelope join at intervals to form "pores" several hundred Å in diameter (Fig 1 5)

How does a "directive" of the nucleus reach the cytoplasm or, conversely, how do cytoplasmic and other external feedback messages reach the nucleus? Non-nuclear substances can act as inducers or repressors (chap 8) of the synthesis of specific proteins in the cytoplasm This almost certainly requires interaction with genes Furthermore, most gene products (e g , mRNA) must leave the nucleus and enter the cytoplasm to express their effects Permeability properties of the nucleus are too complex to be explained by simple holes The morphology of the nuclear boundary provides for two alternative routes for the transfer of information—either across membranes of the perinuclear sac or through "pores " The pores are often referred to as "annuli" to emphasize that they are not simple holes but rather organized regions often pores are seen which contain a diaphragm or plug In addition, the membrane adjacent to the pore may show morphological traces of special organization

To date, the morphological evidence in support of the physiological and biochemical data on transnuclear transport through pores rests mainly on a few observations, such as the movement of electron-dense material (thought to be RNA-containing granules) through the nuclear pores in the insect salivary gland and some other tissues, and the movement of a marker (colloidal gold) into the nucleus when it is injected into the cytoplasm of ameba Clear morphological evidence on passage of material across the nuclear membranes, as distinct from the pores, is not available

THE ENDOPLASMIC RETICULUM

Often an integrated biochemical and ultrastructural investigation (involving cell disrup-

tion, isolation, and analysis of a homogeneous organelle population) leads to the clearest understanding of organelle function in situ From such studies, in a variety of cell types, a fraction of membrane-delimited vesicles, referred to as *microsomes*, is recovered The microsomes perform several functions including the provision of a base for the attachment of ribosomes, the biosynthesis of lipids and, in the case of striated muscle, the accumulation and release of calcium In the intact cell, microsomal vesicles are not found as such, rather one observes a tubular network known as the endoplasmic reticulum (ER) (Figs 1 1 and 1 5) It is assumed that the majority of microsomal vesicles represent a reproducible, preparative artifact arising during cell disruption by the shearing into fragments and closing up of the tubules and sacs of the endoplasmic reticulum

Rough ER. As noted above, the endoplasmic reticulum (and/or the microsomes derived from the ER) can provide a base for the attachment of ribosomes (Fig 1 6) Such ribosome-carrying endoplasmic reticulum is referred to as *rough ER* Microsomal vesicles derived from this rough ER were found to be capable of protein synthesis, the newly synthesized protein appearing in the vesicle lumen In the rough ER in situ the nascent protein likewise can be demonstrated within the reticulum lumen

The rough ER seems to grow by synthesizing more of itself The newly made rough ER may lose its ribosomes, and thus become converted to *smooth ER* The relative proportions of rough and smooth ER vary within different cells for example, the rough ER is extensive in cells which specialize in synthesizing protein for export, while the smooth ER is extensive in steroid secreting cells

Smooth ER. As noted above, the ER which lacks ribosomes is referred to as the smooth ER The membrane of the endoplasmic reticulum carries enzymes which are important in several biosynthetic pathways For example, the enzymes required for the synthesis of steroid are found in microsomal fractions of steroid-secreting cells Enzymes involved in triglyceride synthesis as well as phospholipid synthesis are also found in this fraction, the phospholipid sometimes appearing in the ER as small fat droplets In liver cells, important drug-degrading enzymes are associated with the smooth ER Also, in the hepatocytes the close spatial relationship of the smooth ER with glycogen, the major form of glucose, suggests that the smooth ER may function in glycogen me-

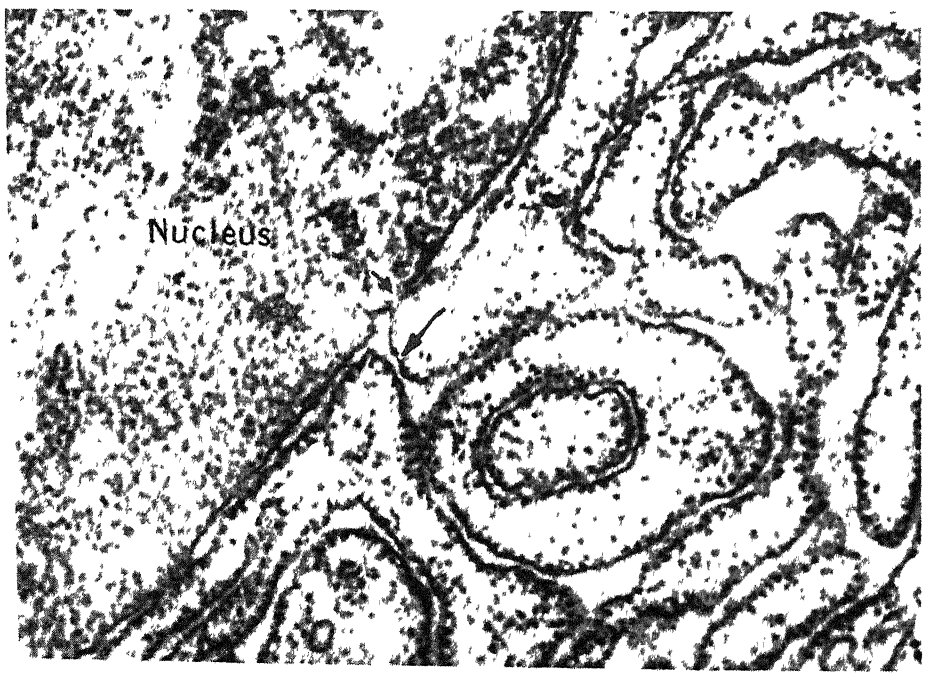


Fig 1 5 Electromicrograph of a portion of a serous cell of mouse salivary gland The rough ER (large arrow) of the cell is continuous with the outer membrane of the nuclear envelope A nuclear pore is indicated by the small arrow (Osmium fixation × 84 000) (Parks from Fawcett, 1966)

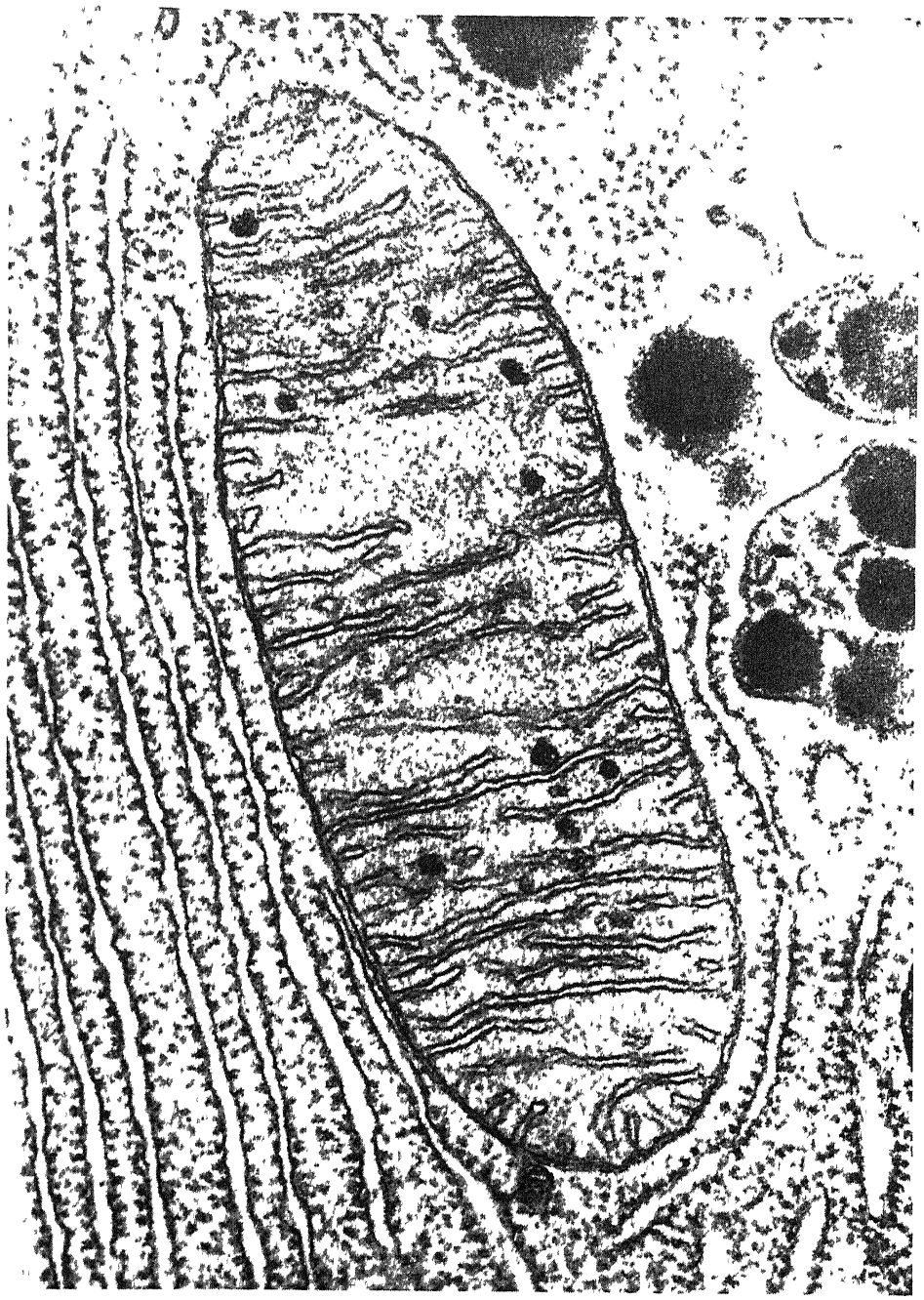


Fig 1 6. Electron micrograph of a portion of a pancreatic cell showing the mitochondrial structures labeled in the diagramatic representation in Figure 1 8 A parallel array of cisternae of rough ER is also evident, 95 000 × (Porter, from Fawcett 1966)

tabolism. In muscle, the smooth ER (sarcoplasmic reticulum) controls the local concentration of calcium ions near the contractile machinery and thereby influences the contraction and relaxation process (chap 4)

THE GOLGI APPARATUS

The Golgi apparatus is believed to be a site for the concentration of protein and polysaccharide. It is also a site for completion of the

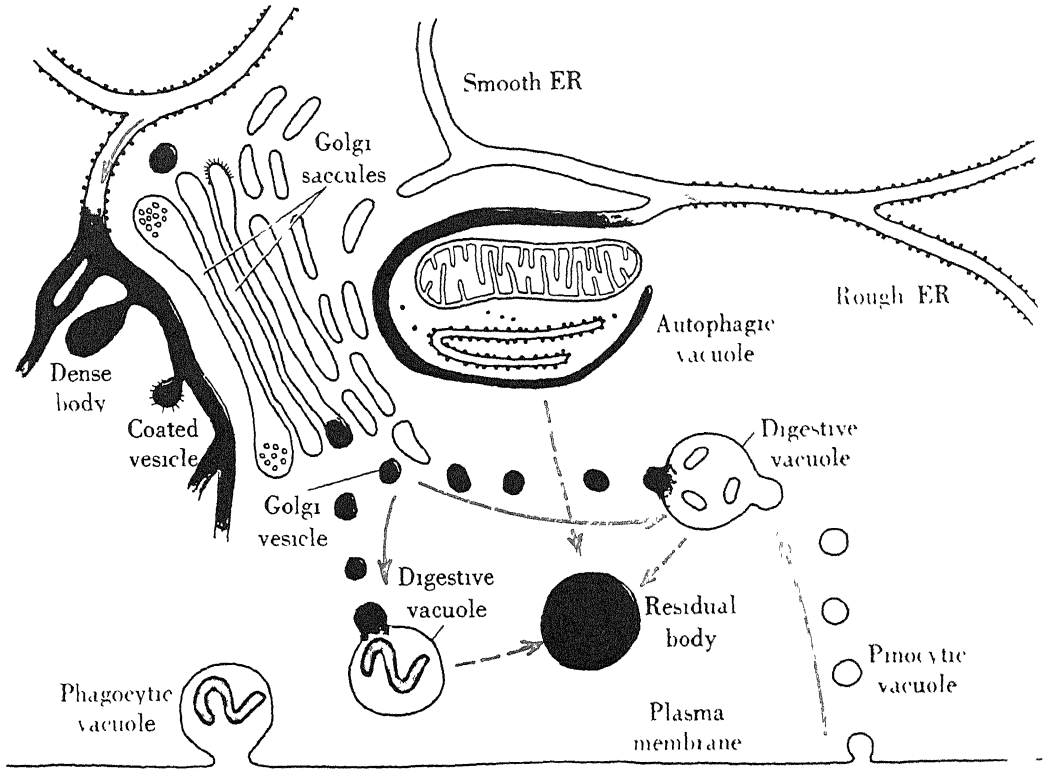


Fig 1 7 Diagram suggesting probable interrelations of some organelles in the transport of lysosomal hydrolases (gray) and the formation of lysosomes Illustrated are *1)* Fusion of primary lysosomes (Golgi vesicles) with phagocytic and pinocytic vesicles resulting in the formation of digestive vacuoles Some of the vesicles are coated (see text) *2)* Movement of hydrolases within the RER and SER to the vicinity of the Golgi apparatus and to lysosomes Dense bodies (bodies containing electron dense grains and other material) and autophagic vacuoles are among the structures that may form from the ER or from closely interrelated components of the ER and Golgi apparatus *3)* Formation of residual bodies by the accumulation of indigestible residues within lysosomes (From Novikoff and Holtzmann 1970)

synthesis of the carbohydrate moiety of glycoprotein, e g , the synthesis of the carbohydrate moieties of thyroglobulin and immunoglobulin begins in the ER, but the terminal sugars are added in the Golgi apparatus These products are usually packaged as "granules" within Golgi-derived vacuoles or vesicles which then migrate away from the Golgi apparatus

The Golgi apparatus consists of a stack of several membranous saccules with associated vacuoles and vesicles (Fig 1 7) The ER in some cell types is assumed to contribute to the "forming face" or "top" of the Golgi apparatus Within the stacked membranes of the Golgi apparatus materials are concentrated as they pass from the saccules on the "top" surface to those forming the "bottom" surface of the apparatus (Fig 1 1) In exocrine and endocrine cells the mature secretory granules are generally found in association with the "bottom" saccules

LYSOSOMES

Lysosomes have been found in virtually all animal cells which have been studied As organelles they are best defined by biochemical and cytochemical criteria a lysosome is a membrane-delimited body containing demonstrable acid-hydrolase activity Over 30 acid hydrolases are known to occur in the lysosomes, these enzymes can digest essentially all macromolecules Material to be digested becomes enclosed within lysosomal membranes permitting isolated, controlled degradation There are numerous findings suggesting that release of hydrolases from the lysosome into the cell may be important in various pathological states,

many of these findings are yet to be fully evaluated

Morphologically, the lysosomes are a motley group of subcellular bodies (Fig 1 7) Their appearance depends largely on the origin of the enclosed material which is destined for intracellular digestion by the lysosomal hydrolases In polymorphonuclear white blood cells Golgi-derived lysosomal granules fuse with phagocytic vacuoles formed as a result of endocytosis of foreign material Other lysosomes are *autophagic*, these lysosomes contain bits of the cell's own substance which have been separated along with the hydrolases from the rest of the cytoplasm within a membrane-delimited space Autophagia, which may be enhanced by stress, is hypothesized to be important for the turnover of some cell constituents

The degraded soluble products of lysosomal hydrolysis can either enter the anabolic pool to be reused in biosynthesis or to be secreted An example of the latter is the secretion of thyroid hormone thyroid colloid travels from the follicle lumen within endocytic vacuoles which then fuse with the lysosomes, the colloid is hydrolysed and thyroxine is released It is also noteworthy that indigestible residues accumulate within lysosomes, a phenomenon which accompanies the aging process in neural and other cells

PEROXISOMES

The peroxisomes constitute another group of membrane-delimited bodies They are often associated with the ER from which they are thought to derive (see Fig 1 1) They are concerned with the metabolism of peroxide peroxisomes contain enzymes which produce hydrogen peroxide (such as amino acid oxidase and in some species, urate oxidase) and other enzymes which destroy hydrogen peroxide (such as catalase) Peroxisomes have, thus far, been found in only a few cell types, of which liver and kidney are among the best established examples The function of peroxisomes is currently under active investigation and there are hints that, in some species, they are involved in carbohydrate synthesis from fat and in the degradation of purines (chap 6) as well as in the detoxification of hydrogen peroxide

MITOCHONDRIA

The early cytologists noted that mitochondria were closely associated with motile processes and were situated in regions of intense metabolic activity such as sites of active transport Biochemists have since shown that two major metabolic pathways, the tricarboxylic acid cycle and the electron transport chain, are situated within the mitochondrion, thus, this organelle is involved in the metabolism of lipids, amino acids, and carbohydrates

The "typical" mitochondrion (Fig 1 6) has a length of 5 to 10 μm and a diameter of 0 5 to 1 0 μm It is bounded externally by two lipoprotein membranes (each about 70 Å thick) the inner one of which is thrown into folds termed cristae or tubules (Figs 1 6 and 1 8) Within the inner membrane is a matrix containing granules, RNA and DNA, the latter is a small circular molecule and is thought to provide some, but not all, of the necessary information for replication of the mitochondria From physiological and biochemical evidence, it is generally believed that the respiratory enzymes and the components of oxidative phosphorylation are associated in an ordered array on the inner membrane The order is thought to promote the sequential interaction of substrates and enzymes in these multienzyme systems with concomitant conservation of energy Therefore, much effort has been directed toward isolation of modular physiological units from inner membrane fractions The fact that multienzyme assemblies can be obtained from mitochondria has been encouraging Moreover, morphological studies support the multienzyme assembly hypothesis First, the respiratory activity of mitochondria is roughly proportional to the amount of inner membrane—a finding which could be explained by presuming an increased number of repeating units or respiratory assemblies associated with the increased membrane area Second, a repetitive array of particles found in certain preparations suggests a similarly repetitive assemblage of inner membrane enzyme systems These regularly spaced particles are found to be attached by small stalks to the inner membrane of isolated mitochondria, unfixed and negatively stained (in negative staining the surface details of the material under study appear as light objects against a dark background) Surprisingly after isolation the stalked particles contained an ATPase predominantly, however, it is suspected that the equilibrium of the reaction catalyzed by the ATPase in the stalked particles is reversed in the intact mitochondrial oxidative

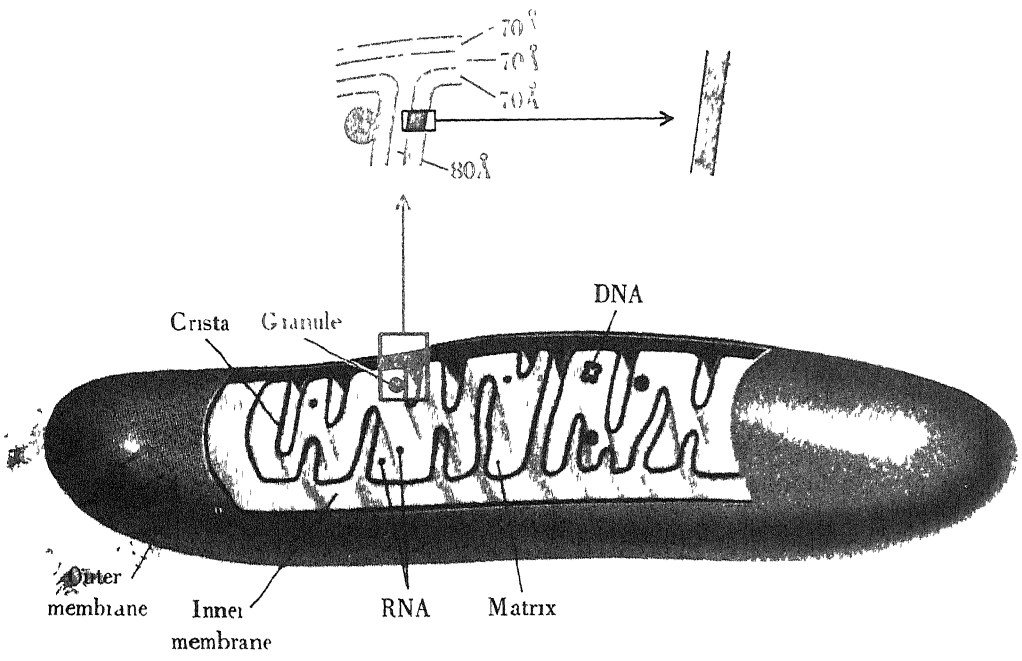


Fig 1 8 Diagrammatic representation of a mitochondrion After work of Ornstein Palade Sjostrand Fernandez Moran and others (From Novikoff and Holtzman 1970)

phosphorylation system, and that it couples phosphorylation of adenosine diphosphate (ADP) with electron transport This suggestion is supported by the fact that dissociation of the particle from submitochondrial preparations abolishes the ability of the preparation to carry out oxidative phosphorylation, when the particles are added back, oxidative phosphorylation returns too These particles are closely associated with the assemblage of enzymes and cytochromes—distributed in the inner membrane—which form the electron transport system of the cell The morphology of the inner membrane enzyme system remains a matter of current dispute

The outer membrane of the mitochondrion has a different enzyme content, a larger percentage of lipids and is more permeable to simple sugars than the inner membrane In addition to the demonstrably membrane-anchored enzymes of both inner and outer membranes, certain enzymes (such as those of the tricarboxylic acid cycle) are solubilized after disruption of the mitochondria These enzymes are presumed to be situated in the matrix or possibly loosely attached to a mitochondrial membrane

Physiologists can clearly define the metabolic state of the mitochondria in terms of electron transport and oxidative phosphorylation Careful electron microscopic study of isolated mitochondria, and in a few cases of mitochondria in situ, has resulted in the identification of two functionally related states 1) the condensed state in which the matrix appears dense and the space between the inner and outer membrane is enlarged, this state is seen when oxidative phosphorylation is proceeding at a rapid rate under conditions of excess ADP and P_i, and 2) the orthodox state (see Fig 1 6) when ADP and P_i are rate limiting in oxidative phosphorylation It is hoped that this kind of structural alteration can be explained eventually in terms of the interaction of mitochondrial macromolecules.

Microtubules and Microfilaments

The asymmetry observed in certain cell types is sharply at variance with the picture of an idealized cell in which all the organelles surround a central nucleus symmetrically A nerve cell in which the axon runs several feet as an extension of the perikaryon, an elongated muscle cell and a squamous (or columnar) epithelial cell all exemplify such asymmetry Likewise the nonrandom (asymmetric) movement of

subcellular elements is exemplified by transport of neurosecretory products (axonal flow), sliding myofilaments in muscle contraction, and chromosomal movement in cell division Electron microscopy of aldehyde-fixed tissue has revealed a morphological basis for asymmetric structure and movement in the form of entities referred to as *microtubules* and *microfilaments*, these structures are not membrane-delimited

MICROTUBULES AND CENTRIOLES

In cross-section microtubules are 200 to 300 A in diameter and may be followed for several microns in longitudinal sections They are found in many regions in which phase-contrast and polarizing microscopy had previously demonstrated the presence of formed oriented, elongated elements Microtubules are often associated with oriented movement The best established example is chromosomal movement in the mitotic spindle the mitotic poles—towards which the microtubules of the spindle orient and towards which the chromosomes move—have centrioles, usually two per cell

In the interphase cell the pair of centrioles are generally found with long axes at right angles to one another Each centriole is a cylinder 0 15 μm in diameter and 0 5 μm in length, composed of nine sets of microtubule-like elements (Fig 1 9)

The organization of microtubules can be disrupted by physical means (freezing or high pressure) or by chemical treatment, especially with colchicine When this is done, motion is inhibited and some of the structure collapses Therefore, the affinity of microtubules for colchicine serves as a means of identifying microtubular protein in a cell fraction Isolated and disrupted microtubules yield protein subunits of approximately 6×10^4 MW These appear to be globular subunits which may be arranged in a helical fashion to form the microtubules In normal cells the microtubular protein appears to be present in a form which is assembled into tubules (e g , for the formation of the mitotic spindle) under appropriate, but as yet not understood, stimulation

CILIA AND FLAGELLA

Microtubule-like structures may also be organized into organelles as diagrammed in Figure 1 9 Cilia and flagella are rapidly beating

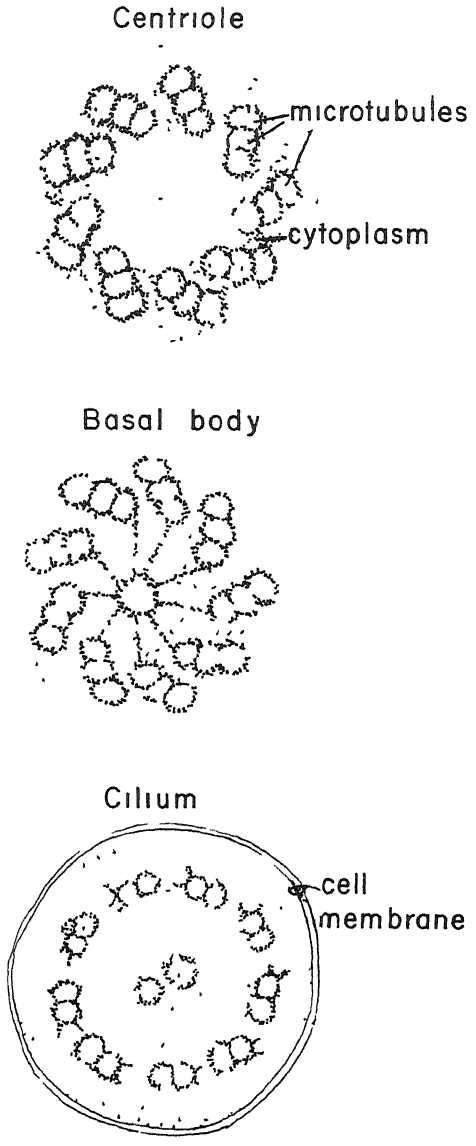


Fig 1 9 Diagrammatic representations of organelles composed of microtubules (After work of Fawcett and Gibbons)

cell processes which extend 10 to 100 μm from the cell The intracellular basal bodies of cilia and flagella are also composed of microtubular structures arranged in the pattern of nine basic units (often referred to as "9 + 0"), they are widely assumed to be an alternate form of the centriole

Cilia and flagella generally have, in addition to the basic nine outer sets, a central pair of microtubules ("9 + 1") It is not known how these elements interact in the matrix to pro-

duce beats, however the process requires cellular energy as indicated by the findings that exogenous ATP can cause beating in isolated cilia and flagella and that the "arms" of the nine sets of microtubules contain an ATPase

MICROFILAMENTS

Microfilaments are a heterogeneous class of long, thin nontubular structures of various diameters Among the most commonly seen microfilaments are those which appear to serve as the structural core of microvilli Also frequently encountered are tonofilaments on the intracellular side of desmosomes (see below) The best example of association of microfilaments with motion is the extensively developed myofilament system which forms the basis of muscle contraction (chap 4)

The Junctional Complex

We have thus far limited our view of the cell to those entities circumscribed by and including its boundary, the plasma membrane Cells rarely are continuous with one another, usually a space of 100 to 200 Å separates them The cells are associated in tissues by various means, the best described is the *junctional complex* in epithelial cells (Fig 1 1) In this complex the plasma membranes of two adjacent cells contribute to specialized attachment sites a tight junction (*zonula occludens*) a desmosome (*macula adherens*) and between these two usually a less well-defined *zonula adherens* where the two membranes are separated by a constant 200 Å space The desmosome may contain organized extracellular material between the two cell membranes which may be seen as an additional dense line in parallel with the membranes

In the region of the tight junction the outermost layers of the two cellular unit membranes appear to be very closely associated or fused with one another (Fig 1 1), externally applied tracer molecules (such as ferritin, an iron-containing electron dense protein) cannot penetrate between the cells at the tight junction Movement across epithelial cell layers with tight junctions requires a pathway through cells rather than around them Between some cells the usual insulation effect on passage of an applied electric current is greatly reduced at the tight junction, such cells are referred to as *electronically coupled* Communication between and coordination of the individual cells of cardiac and smooth muscle may be effected via tight junctions

2 / PHYSICOCHEMICAL PROPERTIES, PERMEABILITY, AND TRANSPORT FUNCTIONS OF CELL MEMBRANES

Several aspects of structure and function of the cell membrane—often referred to as the plasma membrane—were reviewed in Chapter 1 In this chapter we shall further consider the physicochemical properties of this important boundary structure and examine the processes by which it regulates the molecular traffic between the extracellular and intracellular fluids

The two sides of the cell membrane appear to have different properties One may therefore provisionally label the two sides of the membrane as the *outside* (extracellular) and *inside* (cytoplasmic) surfaces In the case of the cell membrane the "outside" surface is in contact with the cell environment, the "inside" surface is in contact with the cytoplasm

Physicochemical Properties

ELECTRICAL PROPERTIES

Even before the nature of the cell membrane had been demonstrated by electron microscopy, electrical measurements gave rise to the interesting generalization that the boundaries of all kinds of cells shared a common property the capacitance per unit area of the cell boundary material was about 1 μf /cm ^{2} This figure can be combined with the 75 Å thickness of the unit membrane in an interesting manner The cell membrane can be pictured, to the first approximation, to be the dielectric material of a charged condenser In general the charge is negative on the inside surface and positive on the outside surface There is a difference of potential across the membrane, for example in resting nerve, on the order of 0 06 v The intensity of an electric field is usually expressed in dimensions of volts per meter Thus, the electric field within the cell membrane of resting nerve is on the order of

$$(0\ 06\ \text{v})/(75 \times 10^{-10}\ \text{m}) = 8{,}000{,}000\ \text{v/m}$$

The dielectric strength of a commercial insulator is the maximum electric field (volts/meter) which the material can stand without breaking down The highest dielectric strength for a good commercial insulator such as rubber is about 1,000,000 v /m Thus, the cell membrane material, viewed as an electrical insulator, must have a very high dielectric strength, i e , 8,000,000 v /m

A second number of interest which we are able to estimate is the dielectric constant, k, of the cell membrane material The capacitance of a condenser is a measure of the amount of charge which can be stored in the plates of the condenser when a given voltage is impressed across these plates The capacitance, C, is directly proportional to the cross-sectional area, A, of the condenser, and to the dielectric constant, k, of the insulator between the condenser plates, the capacitance is inversely proportional to the thickness of the dielectric, x, that is, the distance between the condenser plates In mks units $C = \epsilon_0 kA/x$, where ϵ_0 is the constant 9×10^{-12} farads/m Solving for k and substituting we find that

$$k = \frac{Cx}{\epsilon_0 A} = \frac{(10^{-6}\ \text{farad})(75 \times 10^{-10}\ \text{m})}{(9 \times 10^{12}\ \text{farads/m})(10^{-4}\ \text{m}^2)} \cong 8$$

Again compared to commercial electrical insulators a dielectric constant of 8 is relatively high We conclude that the nerve (cell) membrane has a relatively high dielectric constant and dielectric strength

MOLECULAR PROPERTIES

The postulated molecular structure of the cell membrane is best understood in terms of the hydrophilic and hydrophobic nature of certain chemical groups Hydrophilic groups are polar

groups which may or may not bear a formal electrical charge Substances containing these groups tend to dissolve in water by interacting with the dipolar water molecules, such interactions are enhanced as a consequence of the high dielectric constant of water Accordingly, there is a reduction in the strength of the interactions of polar solutes (dipolar molecules, ions) with each other Amino acids and sugar molecules, for example, are hydrophilic The long-chain hydrocarbons (e g , long-chain fatty acids) on the other hand, are hydrophobic, such molecules are relatively insoluble in water The aggregation of hydrophobic molecules in the presence of water has been explained as follows The structure of water is in an ordered, icelike state in the vicinity of nonpolar molecules When nonpolar molecules leave the water phase and form an aggregate, the degree of order of the water phase is decreased, the entropy (degree of disorder) of the whole system is thereby increased Since systems tend to change spontaneously towards a state in which the entropy is maximized, nonpolar (hydrophobic) molecules tend to separate from the water phase and to aggregate, i e , they do not "dissolve" in water This complex set of ideas for explaining the association of nonpolar molecules is represented in the current literature by the term "hydrophobic bond" (see chap 5)

A long-chain fatty acid is an hydrophobic molecule for the most part but it bears a hydrophilic (carboxyl) group at one end When hydrophobicity exceeds hydrophilicity, such a substance does not appreciably dissolve molecule-by-molecule in water If enough molecules of this kind are present however they can enter the water phase and simultaneously satisfy their hydrophobic and hydrophilic tendencies by forming spherical or lamellar *micelles* A micelle is a multimolecular structure in which the hydrophilic parts of its constituent molecules are at the surface in contact with water, while the hydrophobic parts are in the interior in a milieu of other hydrophobic groups

A considerably simplified model of the cell membrane pictures it to be a lamellar micelle consisting of two rows of phospholipid molecules (Fig 1 10, see also Fig 1 2) The hydrophobic hydrocarbon chains are on the interior The two surfaces of the micelle are covered by a macromolecular material The asymmetry of this membrane is attributed to differences in the macromolecular materials which line the two surfaces of the micelle

The biological role played by the cell membrane can best be interpreted from the point of view of evolution It is unlikely that at the time of the origin of life the first organism was already equipped with a membrane boundary It is more likely that this early organism was a relatively simple, self-replicating molecule With the development of biological complexity, however, an "organelle" arose which isolated the organism from its environment This organelle, the cell membrane, had selective value since it tended to reduce the influx of injurious molecules from the environment while tending to preserve useful molecules for the exclusive use of the organism

From this point of view the cell membrane should be thought of as a barrier to diffusion

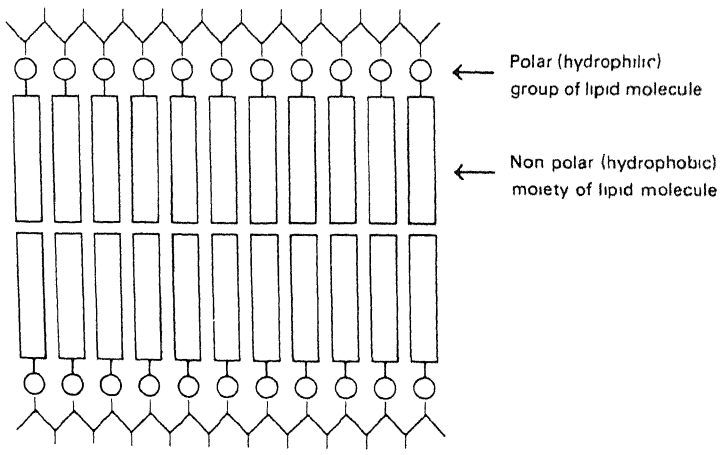


Fig 1 10 Model of cell membrane as laminar micelle

Basically, the cell membrane should be regarded as a relatively impermeable rather than a permeable structure

We have considered the cell membrane to be basically a lipid micelle To the *first* approximation, it is convenient to treat the cell membrane as if it were a thin oil film (Fig 1 11*a*) This crude approximation suffices to explain a number of important permeability and electrical properties of the cell membrane For example, the movement of a substance across the cell membrane may be viewed as follows Each individual molecule of the substance leaves the aqueous phase on one side of the membrane and dissolves in the oil film The molecule then diffuses across the oil film and enters the aqueous phase on the other side

A slightly more sophisticated model of the cell membrane is required to explain a number of experimental findings, which can best be interpreted by assuming that some molecules can pass through the membrane without leaving the aqueous phase In other words, to the *second approximation*, the cell membrane behaves as if it were an oil film pierced by pores (Fig 1 11*b*) The aqueous phase is continuous across the membrane through the pores These pores may be statistical rather than permanent in nature, that is, they may be temporary holes in the oil film created by the thermal agitation From osmotic studies on the human erythrocyte, it has been inferred that the effective diameter of individual pores is about 7 Å and that the total pore area is only about 0 0001 of the total surface area of the cell From the cell size of the pores it is evident that only small particles (e g , water, urea) can cross the cell membrane via the pores Larger molecules will have to cross the membrane in the region which we have referred to as the "oil film" Thus, for the purposes of our survey, it will be useful to view the cell membrane as a perforated oil film To this simple model more complex features (carrier molecules, enzymes, virus attachment sites) can be added as needed

We may now consider the manner and the relative rates at which substances of biological significance cross the cell membrane The following modes of transport will be discussed

1) Bulk Flow In this mode water molecules do not pass through the membrane by dissolving in it and diffusing across it, one by one, rather, the motion of each molecule is correlated with that of neighboring water molecules,

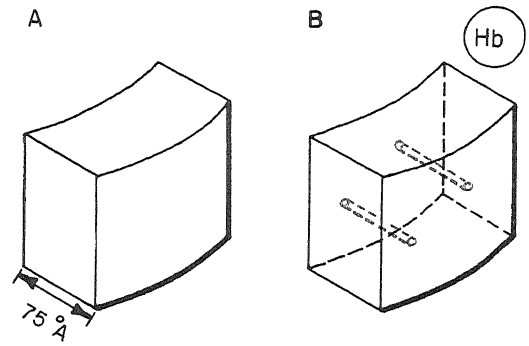


Fig 1 11 Simple models of cell membrane
a) First approximation Oil film 75 Å thick
b) Second approximation 75 Å thick oil film perforated by pores Assuming 7 Å for the average diameter of a pore each pore is about ten pore diameters long The diameter of a hemoglobin molecule (Hb) is shown to the same scale

that is, water flows through the pores in the membrane in bulk as it does through a pipe

2) Diffusion Solute and water molecules cross the membrane individually as a consequence of their thermal motion

3) Carrier-Mediated Diffusion Solute molecules combine with membrane-bound carrier molecules The carrier-solute complex diffuses or otherwise moves across the membrane

4) Active Transport Molecules move across the membrane by a carrier-mediated process which requires a coupled input of metabolic energy

Transport

PHYSICAL BASIS OF DIFFUSION

Consider a tube which is divided into two parts by a thin, removable partition The tube contains an aqueous solution of some substance The concentration of this substance is higher to the right of the partition than to the left After the partition is carefully withdrawn it is found that with time the concentration on the left increases and that on the right decreases Ultimately the concentration becomes the same throughout the tube The physical explanation of this phenomenon is quite simple Every solute molecule has motion of its own and in addition is under constant bombardment from adjacent solute and solvent molecules As a result a given solute molecule follows an erratic path through the liquid However, because there is initially a greater concentration of solute at the right of our tube

than at the left, more solute molecules move from right to left than from left to right until the solute concentration throughout the whole tube is equalized Thus, although each individual solute molecule moves at random, the average motion of a large collection of such molecules results in a net flow of solute from a region of high concentration into a region of low concentration

Diffusion Through a Membrane In view of our interest in the diffusion of molecules through the cell membrane we shall define a number, the *permeability coefficient*, which is descriptive of the ease with which a *given solute* can diffuse through a *given membrane* Clearly this number will be a function both of the nature of the membrane and of the nature of the solute molecule

Consider a tube which is divided into two parts by a fixed membrane of thickness, x (Fig 1 12) The tube is filled with a solution containing a solute, S, which is at concentration $[S]_1$ to the left of the membrane and at concentration $[S]_2$ to the right Assuming that $[S]_1$ is greater than $[S]_2$ there will be a net transfer of solute across the membrane from left to right To simplify the mathematics we assume that over the period of observation the concentrations $[S]_1$ and $[S]_2$ (in moles/cm ^{3}) remain constant with time These restrictions can be approximated in a real system by taking measurements over a sufficiently short period so that the concentrations do not change appreciably, and by stirring the solutions on each side of the membrane Under these conditions there will be a net *rate of flow*, Q (moles/sec) of solute across the membrane This flow rate Q, is proportional to the *concentration gradient*, $([S]_1 - [S]_2)/x$ (When the concentration is equal on both sides, that is when $[S]_1 = [S]_2$ then Q is zero) Q is also proportional to the area A (cm ^{2}), of the membrane, i e , the greater the area the greater the flow rate Introducing the proportionality constant p we obtain the equation

$$Q = pA([S]_1 - [S]_2) \qquad (1)$$

It can be seen that p includes the membranes thickness, x and has the dimensions of centimeters per second

$$p = \frac{Q}{A([S]_1 - [S]_2)}$$

$$\text{(cm)}\left(\frac{\text{moles/sec}}{\text{cm}^2}\right)\left(\frac{1}{\text{moles/cm}^3}\right) = \text{cm/sec} \qquad (2)$$

p is the *permeability coefficient* for the given solute and membrane the greater p the greater the rate with which the solute diffuses across the membrane From equation (2) it can be observed that p defines the flow rate of solute per unit membrane area per unit concentration gradient maintained across the membrane

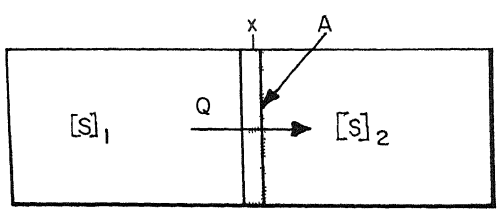


Fig 1 12 Permeability Chamber divided by a thin membrane of area A and thickness x The concentration of a given substance is $[S]_1$ on the left and $[S]_2$ on the right The rate Q in moles/sec at which the substance diffuses from left to right is

$$Q = pA([S]_1 - [S]_2)$$

where p is the permeability coefficient of the membrane for the given substance Note that p may include such factors as the solute diffusion coefficient the membrane thickness the geometry of aqueous channels in the case of polar solutes etc

Any rate of transfer of mass (particles) or energy across a boundary may be referred to as a *flux* Thus, the term Q in equation (1) above may be spoken of as the net flux of solute across the membrane area, A In recent years the letter, J, has been used as a general designation for flux whether the moving particles are molecules, ions, electrons, packets of energy such as calories or quanta, etc In this chapter we are using J to represent the flux, Q, per unit membrane area, A In other words,

$$J \equiv Q/A$$

Therefore, equation (1) for simple diffusion can be rewritten

$$J \equiv Q/A = p([S]_1 - [S]_2) \qquad (3)$$

If the solute particles are ions, the movement or flux can be measured in terms of specific ionic electrical current, I, instead of flux, J Current and flux are related as follows

$$I = zFJ$$

where z is the valence of the ion and F is Faraday's constant (96,500 coul /eq) Typical units

of current are μamp or μamp /cm ^{2} if normalized A handy equation for converting electrical current to ionic flux is

$$I\ (\mu\text{amp /cm}^2) = 26\ 8\ z\ J\ (\mu\text{mole/hr -cm}^2)$$

Physicochemical Criteria for Rates of Diffusion across All Membranes *Lipid Solubility* Let us consider a vessel in which two aqueous phases are separated by an oil phase (analogous to a lipoidal cell membrane) A solute, S, is introduced into aqueous phases 1 and 2 to establish concentrations $[S]_1$ and $[S]_2$, respectively If equilibrium is established across each interface, the concentrations in the oil phase (membrane) and the adjoining aqueous phases can be related to B, the oil-water partition coefficient, which is defined as follows

$$B = [S]_{1\ (oil)}/[S]_{1\ (water)} = [S]_{2\ (oil)}/[S]_{2\ (water)}$$

where $[S]_{1\ (oil)}$ and $[S]_{2\ (oil)}$ are the concentrations of the substance in the oil phases at oil-water interfaces 1 and 2, respectively, and $[S]_{1\ (water)}$ and $[S]_{2\ (water)}$ are the concentrations of the substance in the water phases 1 and 2, respectively There will be a gradient of concentration in the oil phase (membrane) when the concentrations established in the aqueous phases are unequal A gradient across the oil phase gives rise to diffusion of the substance through this phase The net flux per unit area, $J \equiv Q/A$, from aqueous phase 1 to aqueous phase 2 is given by

$$J \equiv Q/A = p'([S]_{1\ (oil)} - [S]_{2\ (oil)})$$

where p' is a proportionality coefficient Knowledge of the oil-water partition coefficient, B, makes it possible to substitute the concentration in the aqueous phases for the concentration in the oil phase at the interfaces (where an equilibrium distribution is assumed to prevail), thus

$$J \equiv Q/A = p'B([S]_{1\ (water)} - [S]_{2\ (water)}) \qquad (4)$$

The apparent permeability coefficient, $p'B$, corresponds to p in equation 3 It is seen that flux through the oil phase (membrane) is proportional to the oil-water partition coefficient, B, all other factors being equal For many substances of biological interest B is very much less than one, for substances which can enter the cell membrane only to a limited extent, Q is small In general, although there are exceptions, the greater the oil-water partition coefficient, B, for a substance, the greater its rate of transfer through the cell membrane (Overton's rule)

Hydrophilicity Chemical groups are said to be polar when the positive nuclei and negative electrons are so unevenly arrayed that the local net electrical charge does not average to zero Let us consider some examples Ionized groups, that is, groups which have lost or gained electrons, are polar Groups in which there has been no net gain or loss of charge, but where a displacement of charge within the group has led to a concentration of negative charge in one region, and positive charge in another, are also polar The alcoholic hydroxyl group is polar because the two valence electrons shared by the hydrogen and oxygen atoms are attracted toward the more electronegative oxygen atom ($-\bar{O}\overset{+}{H}$) This is also why water is a highly polar substance

The greater the number of hydroxyl groups on an organic molecule, for example, the greater the hydrophilicity As one would expect from the lipid (hydrophobic) constitution of biological membranes, the permeability of the cell membrane to a substance is inversely related to the hydrophilic nature of that substance Thus, hydrophilic substances, such as sodium ions, potassium ions, sugars, and amino acids, will penetrate the cell membrane only very slowly by simple diffusion In Table 1 3 it is seen that the permeability constant of sodium ion, Na^+, is only about one hundred millionth that of water It is equally evident, that from the point of view of the economy of the cell certain hydrophilic substances must be brought into the cell in appreciable quantities As we shall see such substances are transported into the cell by a process which involves membrane-bound carrier molecules While it is generally true that electrically charged particles penetrate the cell membrane only very slowly by simple diffusion, it is to be noted that in the case of the mammalian red blood cell, the anions OH^-, Cl^- and HCO_3^- as well as the cation H^+ have permeabilities which are very high compared to that of the cations Na^+ and K^+ In general the electrical charge distribution in the walls of the pores favors the diffusion of anions over cations

Molecular Size In general it may be said for a series of homologous substances that the greater the molecular weight, the smaller the permeability coefficient This follows from the fact that the larger a particle the slower its

thermal motion and, consequently, the smaller its rate of diffusion For polar molecules, such as proteins, nucleic acids, and viruses, the permeability coefficient is exceedingly small not only because of low diffusibility but also because of molecular sieving, i e , large polar molecules tend to be sieved by the cell membrane owing to the geometry (size and shape) and the electrical characteristics (charge distribution) of its aqueous channels Hence, since such particles do indeed cross the cell membrane, we must seek an explanation for their transport in terms of special mechanisms other than simple diffusion

CARRIER-MEDIATED TRANSPORT

The Carrier Model. We have noted that many substances, such as amino acids and sugars, which are vital to the economy of the cell can be expected to cross the cell membrane by simple diffusion only at low rates A considerable body of experimental evidence has accumulated to suggest the following mode of transport The substrate to be transported combines with specific, membrane-bound *carrier* molecules and crosses the cell membrane as part of a substrate-carrier molecular complex The carrier molecules are assumed to be confined to the membrane and to shuttle back and forth from one side of the membrane to the other The carrier is like a ferryboat—it accelerates the rate of transport across a barrier Presumably the carrier reduces the thermal energy needed by the substrate molecule to leave the aqueous phase and to enter the membrane phase Consider a sugar molecule which is strongly bound to the surrounding water by hydrogen bonds For this molecule to enter the nonaqueous phase, the membrane, it is necessary to await a thermal collision of sufficient magnitude to break its hydrogen bonding to the water If, however, there is a carrier molecule in the membrane to which the sugar molecules can readily attach, then the thermal energy needed by the sugar molecules to enter the membrane is decreased because the hydrogen bonds between the sugar and the water molecules can be replaced by bonds between the sugar and the carrier molecules Thus, the entry of the sugar molecules into the membrane is greatly facilitated At the present time many laboratories are engaged in the isolation and characterization of membrane components possessing the required specificity to serve as carriers for the substrate molecules that are believed to undergo carrier-mediated transport in biological systems

Carrier-mediated transport can be distinguished from simple diffusion by the following criteria

Saturation When there is no carrier equation 1 tells us that for simple diffusion the rate of transport is proportional to the concentration difference $[S]_1 - [S]_2$. If this difference doubles, the rate of transport doubles This is not the case for carrier mediated transport Here concentrations can be reached at which virtually all the carrier molecules are in constant use, then further increases in concentration will not appreciably increase the rate of transport since unoccupied carriers are not available The carrier system is saturated and the carrier mediated transport operates at the maximum possible rate

Competition Inhibition Two substances A and B which attach to the same carrier *compete* for the carrier system If B is added to the extracellular fluid, the cellular uptake of A is decreased because molecules of B now occupy carrier molecules which were previously available for the transport of A The competition, of course, is mutual so that addition of A reduces the rate of uptake of B

Specificity One carrier molecule may transport one amino acid or group of amino acids, while another transports one sugar or group of sugars If the addition of a substance X does not decrease the carrier-mediated uptake of a substance A, then A and X do not share the same carrier system and the chemical specificity of the substance-carrier interaction is thereby demonstrated. This specificity is never perfect and a given carrier may carry, with varying efficiencies, a number of substances of related chemical structure

Noncompetitive Inhibition A substance I may interact with a carrier in such a manner as to prevent the carrier from combining with its usual substrate, A In this case, even though I is not being transported, the transport of A is inhibited

Finally, it should be noted that carrier-mediated transport generally shows the high temperature coefficients characteristic of chemical reactions

The carrier-mediated transport of a substance along a concentration gradient (i e , in the direction of simple diffusion) is called *facil-*

ıtated dıffusıon On the level of the organism an analogy to facılıtated dıffusıon may be found ın the transport of oxygen from the lungs to the tıssues The *carrıer* ıs the hemoglobın molecule, the substrate beıng transported ıs oxygen The carrıer manıfests *specıfıcıty*—ıt combınes wıth oxygen but not wıth nıtrogen, *saturatıon*—an ıncrease ın alveolar oxygen tensıon beyond certaın levels does not yıeld a proportıonal ıncrease ın oxygen transport, *competıtıon*—carbon monoxıde competes wıth oxygen for the same carrıer and prevents oxygen uptake

Electrochemıcal Potentıal ($\tilde{\mu}$) The transport processes whıch we have dıscussed so far have dealt wıth the movement of molecules from regıons of hıgh concentratıon to regıons of low concentratıon For generalıty and scıentıfic precısıon we should be referrıng to the movement of molecules from regıons of hıgh *electrochemıcal potentıal* to regıons of low electrochemıcal potentıal However, ın the ınterest of sımplıcıty and clarıty for students who have not been exposed to physıcal chemıstry, we wıll contınue to use the term, concentratıon, ınterchangeably wıth the more rıgorous term, electrochemıcal potentıal Nonetheless, ıt ıs approprıate at thıs tıme to consıder brıefly the meanıng of the electrochemıcal potentıal of a substance, S, ın solutıon ($\tilde{\mu}_S$)

$\tilde{\mu}_s$ ıs a measure of the "useful" energy content of S Typıcal unıts are joule/mole or kcal /mole Electrochemıcal potentıal ıs composed of a chemıcal part and an electrıcal part

$$\tilde{\mu}_s = \underset{\substack{\uparrow \\ \text{chemical} \\ \text{part}}}{\mu_s} + \underset{\substack{\uparrow \\ \text{electrical} \\ \text{part}}}{z_s F \Psi} \tag{5}$$

where z_s ıs the valence of S, F ıs Faraday's constant and Ψ ıs the electrıcal potentıal (voltage) measured wıth respect to any reference poınt The chemıcal part may be separated ınto component parts

$$\mu_s = \mu_s^0(T,P) + RT \ln a \tag{6}$$

where a ıs the actıvıty, R ıs the gas constant, T ıs the absolute temperature, and P ıs the pressure μ_s^0, as ındıcated, ıs a functıon only of T and P, ı e , ıt ıs constant when T and P are constant The actıvıty a may be broken down as follows

$$a = \gamma\ [S] \tag{7}$$

where γ ıs the actıvıty coefficıent (usually ın the neıghborhood of unıty) and $[S]$ ıs the concentratıon of S Therefore, the electrochemıcal potentıal on sıde $ı$ ($ı$ beıng 1 or 2) ıs

$$\tilde{\mu}_{s(ı)} = \mu_s^0(T_ı, P_ı) + RT_ı \ln (\gamma_ı\ [S]_ı) + z_s F \Psi_ı \tag{8}$$

One of the "forces" drıvıng S across the membrane ıs the dıfference ın electrochemıcal potentıal across the membrane In bıologıcal systems, there ıs usually neıther a sıgnıfıcant temperature dıfference ($T_1 = T_2$) nor a sıgnıfıcant pressure dıfference ($P_1 = P_2$) across the membrane and the actıvıty coefficıents are treated as equal ($\gamma_1 = \gamma_2$) In thıs case,

$$\Delta\tilde{\mu}_s = \tilde{\mu}_{s(1)} - \tilde{\mu}_{s(2)} \tag{9}$$

$$= \underset{\substack{\uparrow \\ \text{chemical part}}}{RT \ln ([S]_1/[S]_2)} + \underset{\substack{\uparrow \\ \text{electrical part}}}{z_s F \Delta\psi} \tag{10}$$

where $\Delta\psi = \psi_1 = \psi_2$

ACTIVE TRANSPORT

When a solute moves downhıll from a regıon of hıgh electrochemıcal potentıal to a regıon of low electrochemıcal potentıal, the transport process ıs generally thermodynamıcally spontaneous or dıssıpatıve ın character ın that the free energy of the matter under observatıon decreases durıng transport (Fıg 1 13A), thus, no expenditure of energy ıs requıred on the part of the cell However, a transport process ın whıch the net flow of substance ıs from a regıon of hıgh to a regıon of low electrochemıcal potentıal need not be exclusıvely spontaneous, for example a metabolıc energy source ($\Delta\mu_M$) mıght be coupled to such a downhıll process so as to ıncrease the net rate of flow of substance above that whıch would prevaıl ın a purely dıssıpatıve process A sıtuatıon of thıs type ıs sketched ın Fıgure 1 14 to show the net flux of a substance, J_S^{NET}, beıng drıven by a metabolıc energy source, $\Delta\mu_M$, as well as by the electrochemıcal potentıal gradıent, $\Delta\tilde{\mu}_S$ Thıs ıs an ıllustratıon of *actıve downhıll* transport Another case of actıve downhıll transport ıs observed when the metabolıc component of transport (ı e , that component whıch ıs lınked to $\Delta\mu_M$) ıs orıented ın the reverse dırectıon from but overwhelmed by the electrochemıcal potentıal gradıent, $\Delta\tilde{\mu}_S$ (Fıg 1 15) It ıs also possıble for actıve transport to be manıfest when the electrochemıcal potentıal gradıent has been elımınated ($\Delta\tilde{\mu}_s = 0$) In thıs sıtuatıon, referred to as *actıve level* flow, $\Delta\mu_M$ alone drıves the transport process (Fıg 1 16)

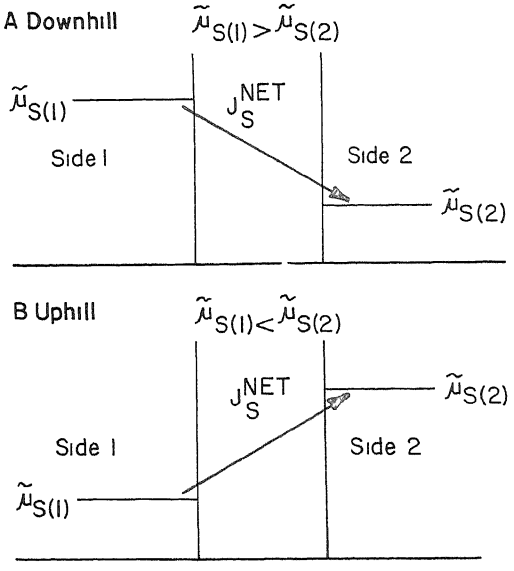


Fig 1 13 If the net flow of a substance is in a direction from higher to lower electrochemical potential it is moving downhill if net flow is from lower to higher electrochemical potential it is moving uphill

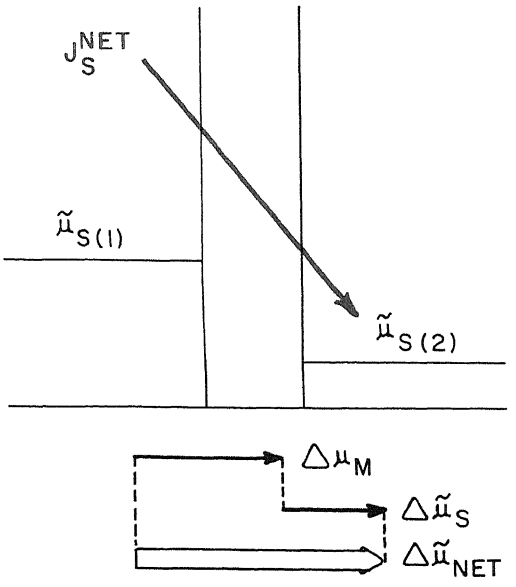


Fig 1 14 Active downhill transport $\Delta\mu_M$ represents the effective input of metabolic energy from the cell for the transport of S This effective input will be less than the free energy available from the coupled metabolic reaction to the extent that the efficiency of coupling between metabolism and transport is less than 100% The metabolic coupling in this scheme is oriented to drive S from side 1 to side 2 $\Delta\tilde{\mu}_S$ is the electrochemical potential difference of S $\Delta\tilde{\mu}_{NET}$ represents the net driving force due to $\Delta\mu_M$ *plus* $\Delta\tilde{\mu}_S$ In this example of active downhill transport both $\Delta\mu_M$ and $\Delta\tilde{\mu}_S$ act to drive S from side 1 to side 2

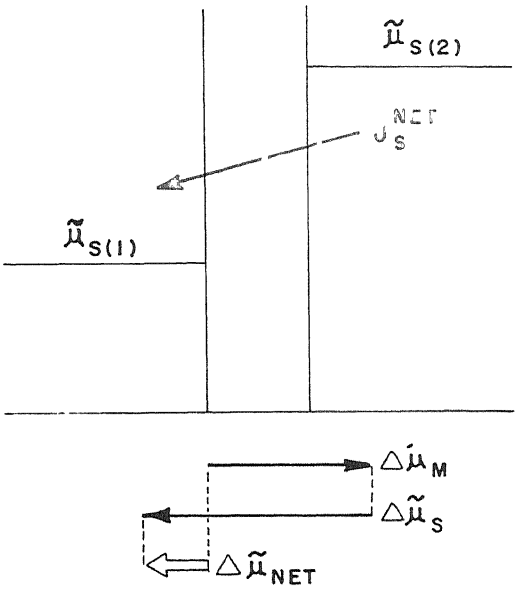


Fig 1 15 Reversed active downhill transport $\Delta\mu_M$ and $\Delta\tilde{\mu}_S$ are oriented to drive S in opposite directions In this example $\Delta\mu_M$ is overwhelmed by $\Delta\mu$ so that $\Delta\tilde{\mu}_{NET}$ is oriented to drive S from side 2 to 1 in the *reverse* of its normal direction of active transport If coupling between the metabolic and transport processes is highly efficient then in principle the reversed downhill transport of S can serve to generate metabolites (e g ATP) from breakdown products (e g ADP *plus* inorganic phosphate)

In all cells there exist mechanisms whereby substances can be transported uphill across membranes from regions of low to regions of high electrochemical potential (Fig 1 13B) These mechanisms are in most cases examples of *active uphill transport* in which a metabolic process (off-equilibrium breakdown of a metabolite) is coupled to and thus provides the energy for the transport process (Fig 1 17) The movement of every mole of substance against its electrochemical potential gradient must be paid for by the expenditure of an amount of energy which depends upon the efficiency of the coupling between the energy-providing process and the transport work accomplished The energy required for active transport is derived from the hydrolysis of substances such as adenosine triphosphate (ATP) If there is no coupling with a metabolic process, transport is passive For example, a substance may move uphill passively if its transport is coupled with the downhill movement of another substance rather than with a metabolic process On the other hand, as we have seen, an

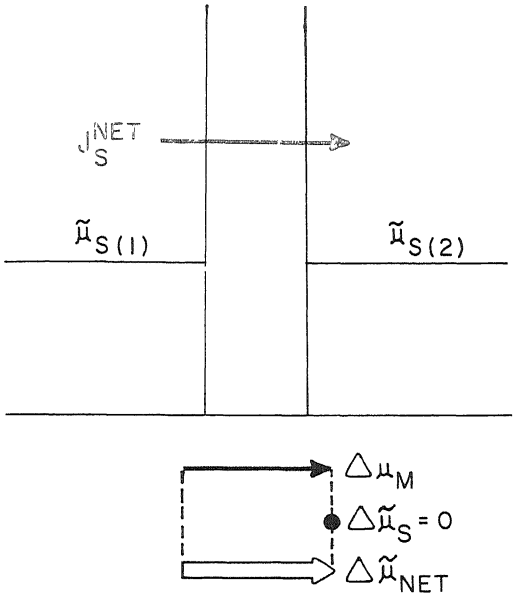


Fig 1 16 Active level flow Only $\Delta\mu_M$ provides the net driving force for the transport of S In undergoing net transport from side 1 to side 2 S neither gains nor loses useful energy ($\tilde{\mu}_S$) The direction of net transport during level flow can be used to define the normal direction of active transport

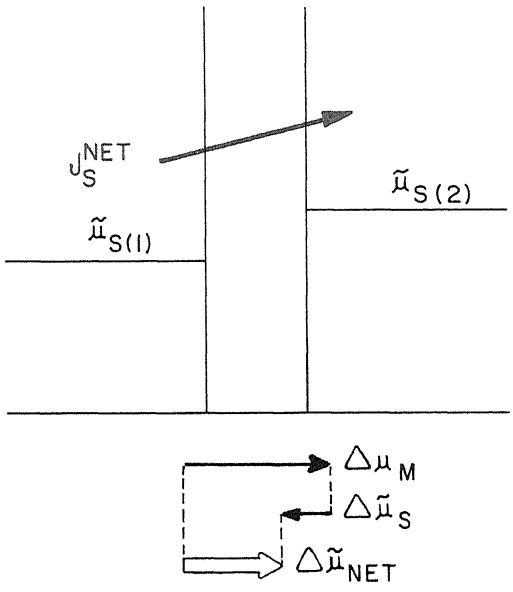


Fig 1 17 Active uphill transport $\Delta\mu_M$ and $\Delta\tilde{\mu}_S$ are oriented to drive S in opposite directions In this example $\Delta\tilde{\mu}_S$ is overwhelmed by $\Delta\mu_M$ so that $\Delta\tilde{\mu}_{NET}$ is oriented to drive S from side 1 to side 2 S gains useful energy during its net transport from side 1 to side 2 In effect the active transport mechanism is responsible for the transfer of energy from metabolites to S

actively transported substance may move uphill, downhill, or remain at a given level with respect to its electrochemical potential, $\tilde{\mu}_S$ Thus, an active transport process is defined, not by demonstrating that flux is thermodynamically uphill, but *only* by demonstrating that flux is coupled to metabolism

In mammalian cells the existence of active transport processes for sodium and potassium ions can be readily inferred from the concentrations of these ions in the intracellular and extracellular fluids (Table 1 1) In general, though with exceptions, the concentration of K^+ is greater in the cell than it is in the extracellular fluid, while the concentration of Na^+ is lower in the cell than it is in the extracellular fluid From the magnitudes of the permeabilities of cell membranes to these ions (Table 1 3) it can be shown that the intra- and extracellular concentrations would equalize by diffusion over a period of several hours It follows that concentration gradients are maintained in vivo across the cell membrane by active processes which pump the ions against the gradients at rates sufficient to offset passive back diffusion (leakage down the gradients)

Active transport is carrier-mediated Therefore the criteria which define it include *1*) saturation, 2) competitive inhibition, *3*) specificity, and *4*) noncompetitive inhibition, however, as noted above, the new criterion which must be added is the requirement for metabolic energy As we have already noted, a substance can be made to move against a concentration gradient only by the expenditure of energy If the energy supply is blocked, active transport ceases As an example we may consider the human erythrocyte which under normal circumstances has a potassium concentration which is higher in the cell than it is in the serum When erythrocytes are refrigerated, active transport ceases and potassium ions slowly leak out of the cells until the intracellular and extracellular potassium concentrations are

Table 1 1
Concentrations (mEq /liter of H_2O) of certain ions in human plasma and erythrocytes

	K^+	Na^+	Ca^{++}	Mg^{++}	Cl	HCO_3
Plasma	5	140	3	1	110	28
Erythrocytes	150	15	Trace	3	70	27
Ratio (*P/E*)	0 03	9	—	0 3	1 6	~1

After Davson, 1964

equal When the blood temperature is returned to 37° C, providing there is an energy source such as glucose in the serum, active transport resumes and the intracellular potassium concentration increases toward normal levels Similar observations have been made in in vitro experiments with muscle tissue, kidney slices, salivary gland slices, etc

A useful model for active transport is shown in Figure 1 18 In addition to the carrier, there is an enzyme, ϵ, and an energy source, W The substance to be transported combines reversibly with a carrier molecule on the extracellular surface of the membrane The substance-carrier complex diffuses to the intracellular side of the cell membrane where the complex dissociates and the substance discharges into the cell In this model the enzyme, ϵ, which is localized on the intracellular side of the membrane, alters the carrier molecule to an inactive form, that is, a form which does not combine with the substance The alteration of the carrier is an enzymatic process which does not require an input of energy The inactivated (unloaded) carrier passes to the extracellular side of the membrane where it is converted to its active form by an energy-requiring process Now it can again combine with the substance to be transported and the cycle then repeats It should be pointed out that there exists no direct evidence for this model The postulated carrier molecule, the inactivating enzyme and the carrier activating system have not yet been isolated or biochemically defined Nevertheless, the model shown in Figure 1 18 is exceedingly helpful in considering a wide range of experimental findings related to active transport It might be thought that active transport should give rise to a continuously increasing concentration gradient This is not the case for the following reasons

1) As the intracellular concentration of the transported substance increases, the dissociation of the substance-carrier complex is opposed by mass action, and therefore, the probability of a loaded carrier molecule returning from the intracellular to the extracellular side increases

2) As the concentration gradient builds up there will be an increase in the passive back diffusion (leak down the gradient in the direction opposite to that of the active transport) so that a steady state is attained in which the rate of active transport equals the rate of back leak

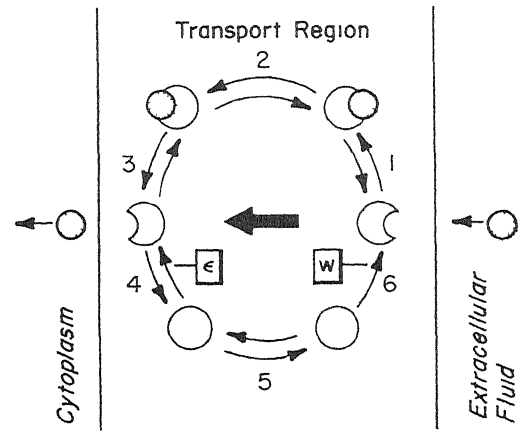


Fig 1 18 Model for active transport The carrier molecules (large circles) are confined to the membrane region There is a net transport of substance (small circles) from the extracellular fluid to the cytoplasm The large arrow indicates the direction of transport

1) Substance combines with carrier

2) Substance-carrier complex passes to cytoplasmic side

3) Substance-carrier complex dissociates Substance discharged into cytoplasm

4) Carrier is enzymatically degraded into an inactive form i e a form which exhibits low affinity for the substance

5) Inactive carrier passes to extracellular side

6) Carrier is activated by an energy requiring process (After Heinz and Walsh, 1958)

Thus, the ratio of the intracellular concentration to the extracellular concentration of the transported substance becomes constant with time

The adaptive value and evolutionary significance of active transport is clear· The total amount of an amino acid in the environment of a microorganism, for example, may be very large though its concentration and rate of entry into the cell may be too small to sustain an adequate level of protein synthesis In the presence of an energy-coupled carrier system the amino acid can be concentrated within the cell. Such a system should not be thought of as operating only between the cell and its outside It can operate at the level of any membrane system—making it possible, for example, for a cell organelle to concentrate a needed constituent This is the case for the mitochondrion in which many ions are accumulated The general features of active transport and facilitated diffusion hold true for a wide range of transport phenomena Active transport processes for amino acids, sugars, and ions (Na^+, K^+, Cl^-,

Mg^{++}, etc) have been demonstrated and referred to by various terms such as the sodium and potassium "pump" of nerve, muscle and erythrocytes, and the active secretion and active reabsorptive processes of the kidney, gastrointestinal tract, and exocrine glands

Unidirectional Fluxes and Net Flux Clearly the net flux across a membrane will be resultant of the two unidirectional fluxes The unidirectional flux of solute, S, from side 1 to 2 ($J_S^{1\to 2}$) can be determined by introducing an isotope S^*, to serve as a tracer for S and then measuring the rate of appearance of S^* on side 2 The unidirectional flux from side 2 to 1 ($J_S^{2\to 1}$) can be determined by introducing a second isotope of S, S^{**}, into side 2 and measuring its rate of appearance on side 1 These operations are depicted schematically in Figure 1 19

Net flux is given by

$$J_S^{i\to j(\text{net})} = J_S^{i\to j} - J_S^{j\to i}$$

If $J_S^{i\to j(\text{net})} > 0$, material S disappears from compartment i and appears in compartment j

Unidirectional flux is of interest because the net flux is often most easily and usefully determined as the difference between the two unidirectional fluxes

Cis- and Trans-. *Cis-* is a prefix referring to the side from which a unidirectional flux originates and *trans-* is a prefix referring to the side to which a unidirectional flux moves

		cis-side	*trans*-side
FLUX	$J_S^{1\to 2}$	1	2
	$J_S^{2\to 1}$	2	1

In general we may represent unidirectional fluxes as follows

cis side *trans*-side

$J_S^{i\to j}$ or $J_S^{cis\to trans}$

It is often important to know whether biologically active agents (drugs, hormones), which modify the transport of a substance, act on one or both unidirectional fluxes and whether such action is from the *cis*-side, the *trans*-side, or both

PINOCYTOSIS

In discussing the relationship between permeability and molecular size it was pointed out that with respect to diffusion the permeability of the cell membrane to large molecules, such as proteins, is negligible Yet the passage of macromolecules into cells is well documented, for example, *1*) the genetic transformation of bacteria by transforming principle (deoxyribonucleic acid), *2*) the infection of cells by viruses, and *3*) the penetration of cells by antibodies One mechanism whereby such macromolecules enter cells is called pinocytosis (cell drinking) a phenomenon related to phagocytosis This phenomenon was noted under the general heading of endocytosis in Chapter 1 In pinocytosis the cell membrane engulfs extracellular fluid and forms a membrane-bounded vacuole (Fig 1 20) The vacuole travels into the interior of the cell The content of the vacuole faces the "outside" surface of the former cell membrane Thus, from the point of view of the asymmetry of the membrane the vacuolar content lies outside the cytoplasm, just as the content of the gastrointestinal tract lies, in reality, outside of an organism However, in pinocytosis, the cell membrane appears to have properties in the cell interior which are different from those which it had on the cell surface The cell membrane of the ameba, for example, is highly impermeable to glucose When pinocytosis is stimulated, however, glucose uptake from the vacuolar fluid becomes considerable It is believed that macromolecules pass from the interior of the pinocytic vacuole into the cytoplasm as a consequence of the breakdown of the vacuolar membrane in the cell interior and the resultant liberation of its contents

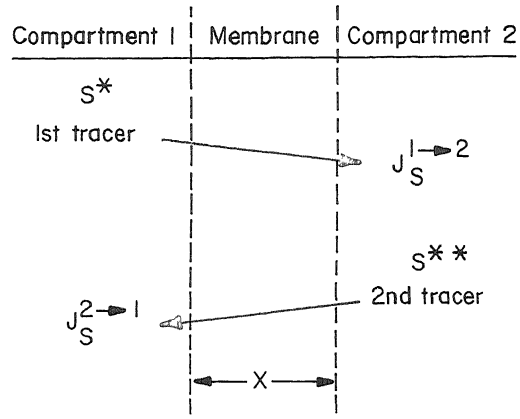


Fig 1 19 Use of tracer isotopes to measure net flux

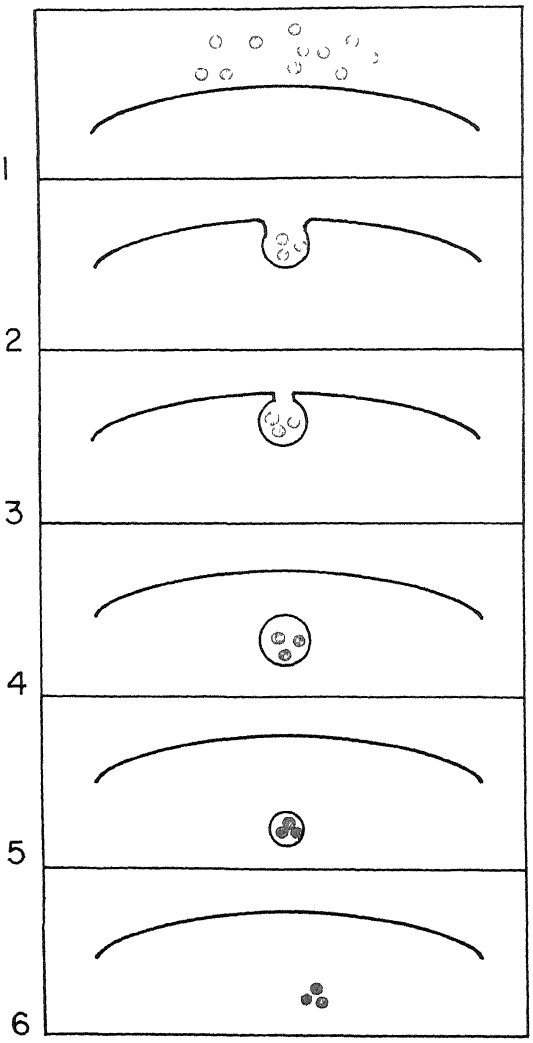


Fig 1 20 Pinocytosis Molecules which stimulate pinocytosis are indicated by dots *1)* Molecules adsorb to the outer surface of the cell membrane *2)* The membrane invaginates in the region of adsorption *3)* A vacuole forms *4)* The vacuole travels inwards *5)* During its inward journey the vacuole shrinks *6)* The membrane disappears and the adsorbed molecules spill into the cytoplasm (After Bennett 1956)

Pinocytosis is not simply the entrapment of a volume of extracellular fluid with the subsequent uptake of solutes from the fluid In the ameba the concentration of protein in the pinocytic vacuole exceeds that in the extracellular fluid This anomaly was explained when it was shown that the protein was adsorbed to the cell membrane prior to the formation of the pinocytic vacuole. It, thus, appears likely that the transport process involved in pinocytosis is primarily one of plasma membrane movement into the interior of the cell and therefore that the ingestion of the extracellular fluid which is trapped in the vacuole is a by product of the process Thus, pinocytosis might be regarded as a transport process in which the carrier is the cell membrane

PHYLOGENETIC VARIABILITY

Thus far we have emphasized the similarity of various biological membranes It is appropriate now to take note of the great diversity of properties manifested by biological membranes In Table 1 2 we note a 100 fold difference in permeability to water of the cell membranes of *Ameba proteus* and the human erythrocyte Even comparisons within the class of erythrocytes show marked differences The permeabilities to glycerol of the cell membranes of erythrocytes of two evolutionary cousins the rat and the ox, differ by a factor of more than 100 We have already considered several generalizations relating the chemical structure of a solute molecule to its permeability coefficient These generalizations were derived largely from studies on plant cells but, with many exceptions, apply also to mammalian cells

Movement of Water Across the Cell Membrane

Because of the special role which water plays in biological systems and because of the unique physical characteristics of this molecule we shall treat the diffusion of water across the cell membrane as a special case Compared to other molecules water diffuses across the cell membrane at a very rapid rate The permeability coefficient of water is 400 times greater than that of urea (Table 1 3) although the latter molecule itself has a very high permeability constant as compared to other molecules of biological significance As a result the water within an erythrocyte exchanges with water from outside the erythrocyte in a small fraction of one second Nevertheless it is important to recognize that the cell membrane is a rather impermeable structure as demonstrated by the fact that a water layer of equal thickness would have a permeability to water molecules 10,000 to 100,000 times larger Thus, when we say that the cell membrane is highly permeable to water we are speaking only in relative terms, that is,

Table 1 2
Permeability to water of some cell membranes

Species	Extracellular Fluid	Permeability (μ^3 water/ μ^2 surface area/atm)
Ameba proteus	Fresh water	0 03
Arbacia egg	Sea water	0 4
Human erythrocyte	Serum	3 0

Prosser and Brown, 1961

Table 1 3
Orders of magnitude of the permeability coefficients of erythrocyte membranes

Substance	Oil/Water Partition Coefficient	Permeability Coefficient (*cm /sec*)	Species
Water	10^{1}	10^{-2}	Ox
Urea	10^{-4}	10^{-4}	Ox
Cl^-	—	10^{-4}	Man
K^+	-	10^{-8}	Man
Na^+	—	10^{-10}	Man

For references see Davson, 1964

we are noting the permeability of the membrane to water as compared with its permeability to other molecules

Osmotic phenomena arise as a consequence of the high rate, as compared to other cell constituents with which water crosses the cell membrane Insight into osmotic phenomena can be most easily obtained by studying a comparable system of deal gases

Consider a container (Fig 1 21*a*) divided into two equal parts by a partition There are an equal number of ideal gas molecules on each side of the partition What distinguishes the two sides is that on the left there are 12 big molecules (marbles) and 4 small molecules (dots), a total of 16 molecules, while on the right there are 16 small molecules (In the biological system the dots correspond to water molecules, the marbles to solute molecules which penetrate the cell membrane very slowly, if at all) Now it is a characteristic of the ideal gas that its pressure depends only on the concentration of particles and *not* on the *nature* (large or small molecules) of the gas Hence, the pressure is the same on each side of the partition Let us now drill holes through the partition of a size that the dots can pass through but the marbles cannot (Fig 1 21*b*)

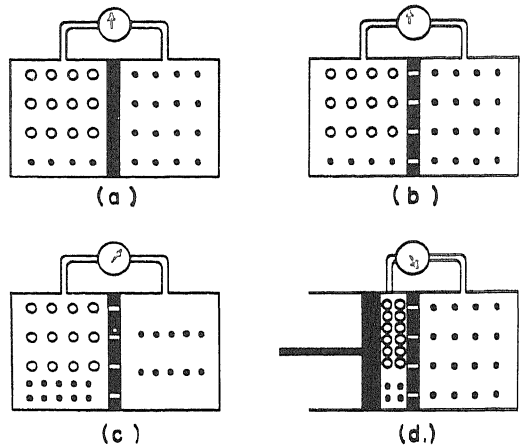


Fig 1 21 Ideal gas model of osmosis

a) An equal number of particles on each side of the partition leads to a zero pressure difference between the two compartments

b) Holes are drilled through the partition These are of such size that the dots but not the marbles can pass from one compartment to the next

c) At equilibrium the concentration of dots is the same on both sides of the partition Due to the presence of the marbles the pressure is higher on the left than it is on the right

d) The net flow of dots from right to left is averted by decreasing the volume of the left compartment and thereby increasing the concentration of dots on the left The pressure difference between the two compartments when the net flow of dots is zero is the osmotic pressure π

What happens next is a consequence of the fact that in a mixture of ideal gases each component behaves as if the other components did not exist Thus, dots migrate through the partition until the number of dots on the left equals the number of dots on the right (Fig 1 21*c*) More generally, we can say that the migration will occur in a direction to equalize the number of dots per unit volume (concentration on both sides of the partition) The migration can be prevented by decreasing the volume on the left side of the chamber so as to equalize the initial concentration of dots on both sides of the *semipermeable* partition (Fig 1 21*d*) The decrease in volume is accompanied by an increase in the pressure of the gas in the left chamber This increase in pressure, π, required to prevent the new flow of "solvent" (dots) between the two chambers is analogous to the *osmotic pressure* of a "solution" (marbles dissolved in dots) with respect to a pure "solvent" (dots)

To transpose this problem back to the cell: the perforated partition corresponds (in its osmotic properties) to the semipermeable cell membrane. (A semipermeable membrane is one which cannot be penetrated by solute molecules or which is very much more readily penetrated by solvent than by solute molecules). The small particles (dots) correspond to the solvent (water), the large particles (marbles) to the solute. As compared to water the marbles represent solute molecules which penetrate the cell membrane slowly, if at all. For an aqueous solution the osmotic pressure, π (in atmospheres), can be estimated from the equation.

$$\pi V = n_0 RT.$$

In this equation V is the volume (in liters) of the solution, n_0 is the number of moles of solute particles in this volume, T is the absolute temperature (°K) and R is the universal gas constant 8.2×10^{-2} (liter·atm.)/(mole·°K). If the extracellular and intracellular fluids have the same osmotic pressure and no solute penetrates the cell membrane, then there is no *net* flow of water across the cell membrane and the cell does not swell or shrink. Mammalian cells do not swell or shrink when placed in a 0.9% NaCl solution. The molecular weight of NaCl is 58.5; therefore the molarity of this solution is 0.15. However, the NaCl is almost completely ionized so that there are approximately 2 osmotically active particles (Na^+ and Cl^-) for every NaCl molecule. Hence, $n_0 = 2 \times 0.15 = 0.3$ moles (per liter), V = 1 liter, T = 310° (at body temperature) and the osmotic pressure of the interior of the mammalian cell is approximately

$$\pi = n_0 RT/V = (0.3)(8.2 \times 10^{-2})(3.1 \times 10^2)/1 = 7.5 \text{ atm.}$$

Any system, such as the cell or the mitochondrion, which is bounded by a semipermeable membrane manifests osmotic properties. In the mammalian organism the osmotic pressure of the blood and extracellular fluids is regulated very precisely and water equilibrates rapidly across the cell membranes. In the ameba, a fresh water organism, there is a constant influx of water into the organism and a continuing energy expenditure on the part of this organism is required to pump the water out again. Thus, the relatively low permeability to water of the ameba as compared with the human erythrocyte (Table 1.2) is an important adaptation for this organism.

3 / EXCITATION, CONDUCTION, AND TRANSMISSION OF THE NERVE IMPULSE

The nervous system is a complex array of specialized structures which serve to receive, store, and transmit information—thereby integrating the activities of spatially separated cells tissues, and organs and making it possible for a multicellular organism to function as a coordinated unit in terms of growth, development, the ability to do work, and to adapt to changes in the environment Our efforts to understand the function of the nervous system involve the concepts and languages of many disciplines ranging from mathematics and physics to the physiology of human beings in all of their complexity In this chapter, however, we will be concerned primarily with the generation and propagation of the impulses which constitute the currency of all neural transactions

The Structure of Nervous Tissue

THE NEURON

The neuron consists of a body (soma, perikaryon), and two types of process—the dendrite and the axon (axis-cylinder process, Fig 1 22A In vertebrates, the bodies of the nerve cells lie within the gray matter of the central nervous system or in outlying ganglia, e g , posterior spinal root, cranial, or autonomic ganglia The white matter of the brain and spinal cord and of the peripheral nerves is composed of bundles of nerve fibers. The core of each nerve fiber is formed by a process of a nerve cell, and many of them are surrounded by a sheath of myelin which gives them a white appearance The gray matter receives a rich blood supply from the vessels of the pia mater; the blood supply to the white substance is much less profuse

There are a number of different types of nerve cell; those in which axon and dendrite arise by a common stem are called unipolar, and those in which the axon and the dendrite or dendrites spring from opposite or at least different parts of the soma are called bipolar or multipolar The cell bodies or somata are of various sizes and forms—stellate, round, pyramidal, fusiform, etc

After fixation and staining by special techniques, various structures are seen in the cytoplasm or perikaryon of the nerve cell body *1)* neurofibrils, *2)* Nissl bodies or tigroid substance, *3)* Golgi apparatus, *4)* mitochondria, *5)* ribosomes, and *6)* the endoplasmic reticulum (see also chap 1) Electron microscopy has revealed much of the detailed structure of these intracellular entities Their structure and function appear to be the same in all cells studied so far Mitochondria, found along the entire length of the axis cylinder, contain all the enzymes required for the respiratory activity of the cell, and are therefore, responsible for those functions dependent upon aerobic metabolism They also participate in protein synthesis but the relationship of this to neural function is not yet clear The neurofibrils appear as fine filaments which stream through the cytoplasm from dendrites to axon (Fig 1 22B), they enter the latter process and extend to its terminations The Nissl bodies composed of ribonucleic acid (RNA) and polysomes are granular masses stainable with basic dyes and occur in the perikaryon and dendrites but not the axon They give a striped or tigroid appearance to the cell They are absent from the region of origin of the axon (axon-hillock) and vary in size and number with the state of the neuron, they undergo disintegration (chromatolysis) in a fatigued or injured cell or in one whose axon has been sectioned This means that the synthetic machinery of the perikaryon is at least partially regulated by events in the peripheral processes of the cell The nature of this feedback control is not known The internal reticular apparatus of Golgi is a coarse network seen within the cells when special methods—e g , impregnation with silver chromate—are employed which leave the Nissl bodies and the neurofibrils invisible

The nucleus of the nerve cell contains one and sometimes two nucleoli but, as a rule, no centrosome The absence of a centrosome indicates that the highly specialized nerve cell has lost its power of division Nerve cells once destroyed are replaced merely by neuroglia

The linking together of neurons to form conducting pathways is effected by the contact (but not union) of the axon terminal of one nerve cell with the body, dendritic process or in some instances, the axon of another Such a junction, without anatomical continuity is called a *synapse*

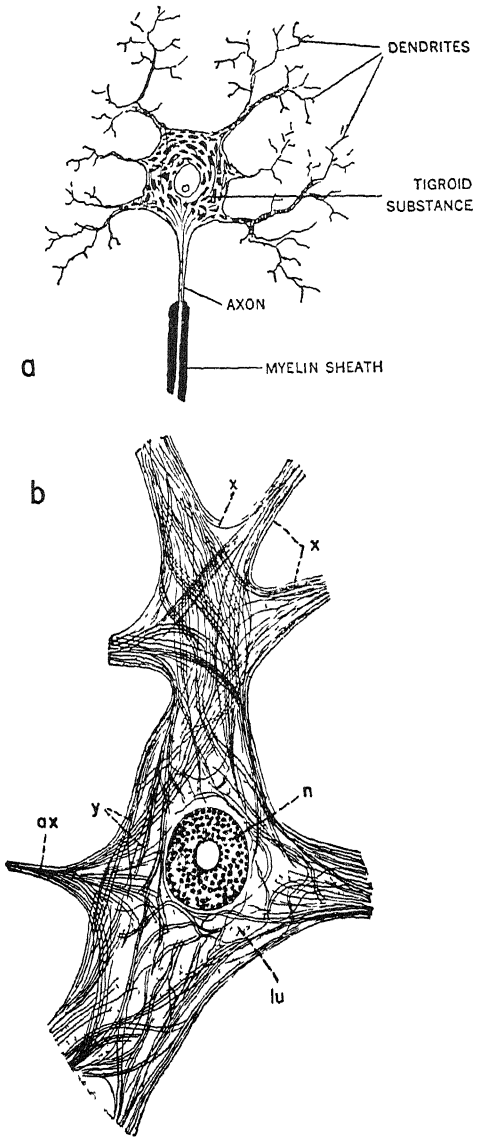


Fig 1 22 **A** Showing different parts of the neuron *B* Showing neurofibrils in a cell from the anterior gray column of the human spinal cord *ax* axon *lu* interfibrillar spaces *n* nucleus *x* neurofibrils passing from one dendrite to another *y* neurofibrils passing through the body of the cell (from Ranson after Bethe and Heidenhain)

Though the nerve cell frequently possesses more than one dendrite the axon is single The axon may be long and contribute to one of the tracts of the central nervous system forming the white matter, or it may terminate as a peripheral nerve fiber Such cells are referred to as Golgi I type In the Golgi II type cell, the axon is short and ends within the gray matter by making contact with another neuron The axon arises from a small elevation on the surface of the cell body, the axon hillock It may give off short collateral branches or run as an unbranched fiber, not dividing until it has reached its destination The dendrite is the receptive process of the neuron, the axon is the discharging process, i e , the former transmits the impulse toward the cell body and the latter away from the cell body Within the central nervous system the dendrite is usually short and possesses many branches, but in the peripheral sensory nerves it is comparable in length with an axon Nerve fibers which carry impulses to the central nervous system are termed *afferent*, those conveying impulses from the central nervous system to the periphery are called *efferent* A mixed nerve contains fibers of both types The fact that conduction is unidirectional is a consequence of the properties of synapses, for nerve fibers per se can be made to conduct in either direction

As with other cells, the cytoplasm of the soma is enveloped by a plasma or unit membrane composed of lipoprotein, a bimolecular leaflet of lipid material covered by a layer of protein This extends over the processes of the nerve cell The electrical properties of nerve depend upon the plasma membrane

THE NERVE FIBER

The white matter of the central nervous system and the peripheral nerves are composed of thousands of individual nerve fibers Within the gray matter the axis cylinders are enclosed only by the plasma membrane; but upon leaving the gray substance they acquire a sheath of lipid material called myelin. The myelin sheath is associated with somatic nerves of large diameter and generally not with any

nerve fibers of small diameter Hence, we speak of myelinated (medullated) or unmyelinated (unmedullated) nerves. Myelinated fibers, usually larger than 1 μ in diameter, conduct faster than the smaller unmyelinated fibers

The myelin sheath of the somatic nerves consists of compressed layers of Schwann cells spiraling concentrically to form a wrapping around the axons The outer layer of the myelin contains the flattened nuclei of the Schwann cells, and is termed the neurilemma or sheath of Schwann Myelinated nerves appear as if constricted at regular intervals along their course This appearance is due to the absence of myelin at these points and the dipping inwards of the neurilemma they are known as the nodes of Ranvier The segments between the nodes vary in length in different nerves, but the usual internodal distance is about 1 mm Each internodal segment of the neurilemma consists of a single Schwann cell The fibers forming the white matter of the central nervous system and optic nerves have no neurilemma, this membrane is replaced by glial cells which produce the myelin sheaths as the Schwann cells do peripherally

Compared with the myelin sheath the neurilemma in the region of the node of Ranvier is highly permeable to ions There is good evidence that the conduction of the impulse along a myelinated nerve is a "leaping" from node to node over the intersegmental regions rather than a continuous process This type of conduction is called saltatory (from the Latin word *saltus*, a leaping)

Myelination of Fiber Tracts in the Central Nervous System The nerve fibers in the various conducting pathways receive their myelin sheaths at different ages and it is generally believed that the myelination of a given tract and the time at which it commences to function coincide. The sensory tracts become myelinated first, those of the posterior columns of the spinal cord between the 4th and 5th months of fetal life (human) The spinocerebellar tracts are myelinated later and the motor paths, e g , corticospinal (pyramidal) tracts, do not begin to be invested with myelin sheaths until the 2nd month of life and are not completely myelinated until about the 2nd year, about the time when the child has learned to walk. The fibers of association paths, for the most part, myelinate at still later dates The function of the myelin sheath is not known, but the high insulating property of myelin suggests that it acts to confine the nerve impulse to individual fibers and thus to prevent cross stimulation of adjacent axons

The neuroglia or "glia" (Greek, glue) is a special type of interstitial tissue Its cells are of three kinds, astrocytes, oligodendrocytes and microglia The microglia appear to have a phagocytic role since they wander into the central nervous system from the meninges and blood vessels, and increase in number during inflammatory processes The astrocytes are found surrounding blood vessels and are thought by some to make up the blood-brain barrier Evidence has been obtained that the oligodendroglia are involved in the formation of the myelin layers around axons within the central nervous system Since it has been shown that the extracellular space of the central nervous system is very small, the glia appear to have permeability properties which enables them to function as a special kind of extracellular "space "

DEGENERATION AND REGENERATION OF NERVE

When a peripheral nerve is cut, the part of the nerve separated from the cell body shows a series of chemical and physical degenerative changes At the same time the fibers of the proximal stump of the nerve, those still attached to their cell bodies, grow distally toward the separated part of the nerve, these changes constitute the process of regeneration (Fig 1 23)

Degeneration. The degenerative period may be divided into an early and late phase Shortly after the nerve has been sectioned, the axis cylinder swells, the myelin sheath begins to form beadlike structures and a series of enlargements (round, fatty fragments) appear)The axon breaks up and its parts are devoured by macrophages. Up to 3 days after section, the distal nerve will continue to conduct an impulse. Changes in the action potential can be observed as early as 2 days after section After the 3rd day, the ability of the nerve to conduct has seriously deteriorated and after the 5th day, an impulse can no longer be evoked.)The first period (3 to 5 days) is one in which the obvious changes in nerve are largely structural Changes in the ultrastructure of the myelin sheath can be shown to occur during this period, as well as

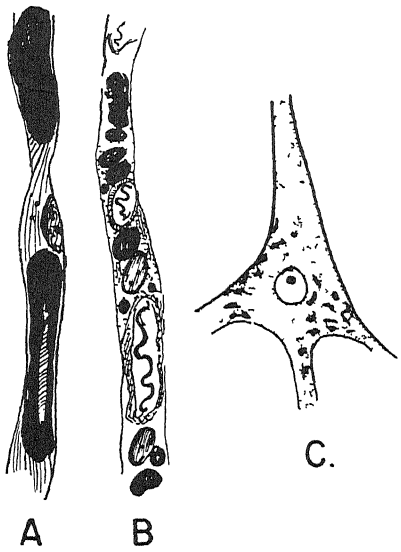


Fig 1 23 Degenerating nerve stained with osmic acid *A* shows appearance of distal segment of nerve fiber 2 days after section note large masses of myelin derived from medullary sheath *B* 5 days after section smaller myelin particles together with droplets of fatty acids and fragmented neurofibrils *C* retrograde degeneration in cell body disintegration of Nissl bodies

changes in the endoplasmic reticulum, mitochondria, neurofibrils, and the plasma membrane

Up to the 8th day, little or no changes in lipid histochemistry may be detected, although changes in nerve cholinesterase and failure in the ability to synthesize acetylcholine have been demonstrated about the 3rd day The lipids appear to maintain their original chemical structure From the 8th to the 32nd day after section, the myelin gradually disappears, while Schwann cells and macrophages have both increased greatly in number Why the myelin breaks down is not clear It has been suggested that the macrophages and Schwann cells secrete enzymes which aid in the destruction of myelin The principal myelin lipids are free cholesterol, and two lipids containing sphingosine cerebroside and sphingomyelin) At the time of the disappearance of the myelin, cholesterol esters appear in large quantities and free cholesterol disappears

The changes just described are generally known as Wallerian degeneration In addition there are other changes that occur in the neuron on the proximal side of the section (retrograde degeneration) The nerve fiber as far centrally as the first node of Ranvier shows changes similar in nature to those just described In the cell body itself swelling of the cytoplasm and nucleus occurs and the Nissl granules undergo disintegration (chromatolysis) Atrophy of the cell body may ultimately result

Regeneration. Following section of a nerve, the fibers in the central stump begin to send out branches consisting of outgrowths of the axon near the cut tip Up to 50 branches may sprout from a cut axon At the same time, there is a rapid proliferation of the Schwann cells If a gap larger than 3 mm exists between the central and peripheral stumps, the fibers tend to intermesh and form a tumor-like swelling called a neuroma In such an event regeneration will probably never occur For this reason it is necessary to join accurately by suture the proximal and peripheral stumps If the neuroma is composed of sensory fibers it may be very painful to pressure, often a troublesome complication following amputation When peripheral degeneration has proceeded for enough time so that the peripheral stump contains only empty neurilemmal tubes, one of the outgrowing sprouts enters a tube and grows to form a new axon Only one of the many sprouts enters an empty tube The ability of axons to grow into a peripheral stump has led some observers to invoke a special chemical attraction between the nerve fiber and the terminal organ This is the doctrine of neurotropism No direct evidence exists for such an attractive force

The rate of growth of regenerating nerve is from 1 to 4 mm /day It has been established, in a variety of experiments, that there is a flow of axoplasm down the normal nerve at about the same rate This axoplasmic flow is accompanied by the movement of protein and nucleoprotein However, it is not known what role axoplasmic flow has in the regeneration of nerve

Principles of Bioelectricity

It will be recalled that the practical unit for quantitation of charge is the coulomb which equals $6\,2 \times 10^{17}$ electrons. A positive charge results from the removal of electrons from a neutral body, as when one electron is removed from the neutral sodium atom to form a sodium ion

An electric field exists in the space around a charge and extends to infinity The strength of this field is measured by the force which would

be exerted on a unit positive charge placed at a point in the field The intensity of the field decreases inversely as the square of the distance from the charge (Coulomb's Law) In order to move a charge in the field, work must be performed against the field Thus, it takes work to move a positive charge up to a region where another positive charge is situated and the potential at any point in the field is defined as the work that must be done on a unit positive charge to move it up to that point from an infinite distance This is the absolute definition of potential which is seldom necessary in practice The significant quantity in experimental work is the difference in potential between two points which is defined as the work necessary to move a unit charge from one point to the other The work is measured in joules The practical unit of potential difference is the volt, i e the potential difference against which 1 joule of work is done in the transfer of 1 coulomb

THE ORIGIN OF POTENTIALS

A difference in potential is associated with a flow of current in a medium (e g , electrons in metal conductors and charge-carrying ions in solutions of electrolytes) The current flow requires energy since it encounters resistance in the medium The relationship between E, the potential difference (in volts), I, the current (in amperes which is the same as coulombs per second) and R, the resistance (in ohms), is given by Ohm's law as follows, $E = IR$ In order to produce electrical currents, special sources of electrical energy are available such as the battery and generator Such sources of energy are said to produce an electromotive force or emf In solutions, electrical energy is produced by chemical reactions which result in the separation of charge at electrodes Other sources of energy produce separation of charge by the flow of ions from solutions of high concentration into solutions of lower concentration, such separation of charge itself constituting a form of stored energy which is manifest as a difference in electrical potential

Two sources of energy are important in understanding the origin of biological potentials, these sources give rise to emfs called *electrode* or *concentration* potentials and *diffusion* or *membrane* potentials. Concentration potentials arise in every measurement of potential difference in solution and are generally to be avoided in biological measurements They arise whenever an electrode consisting of an electrolytic salt is dipped into a solution containing an ion in common with the electrode Thus, an electrode made of silver and coated with silver chloride inserted into a solution of sodium chloride will produce a potential difference between the electrode and the solution The silver dissolves in the solution as silver ion, Ag^+, leaving the electrode negative The positive Ag^+ ion remains at the electrode, forming a double layer of charge with the negative electrode Charge is therefore separated at the electrodes If two such electrodes are set up, each dipping into a different concentration of sodium chloride (a concentration cell) double layers will be set up at each electrode, but the double layer at one electrode will be more densely charged When two such electrodes are connected by a wire, a current will flow The emf produced between the two electrodes can be calculated from the Nernst equation for the concentration cell,

$$E = 2\,303\,(RT/F)\log\,(C_1/C_2)$$

where the electrolyte is univalent R is the gas constant (1 99 cal /mole/°C), T the absolute temperature (°K), F the Faraday, (23,050 cal /v /mole, 96,494 coul /mole), C_1 and C_2 are the concentrations of electrolyte in each solution Note, that according to the formula if the concentrations are equal in each solution, the emf is zero The method of producing a concentration cell is shown in Figure 1 24

Diffusion Potentials. Let us place two solutions containing different amounts of NaCl in contact by means of membrane freely permeable to the electrolyte An emf whose origin is quite different from that of the electrode potentials we have discussed will arise at the membrane The potential difference arises from the

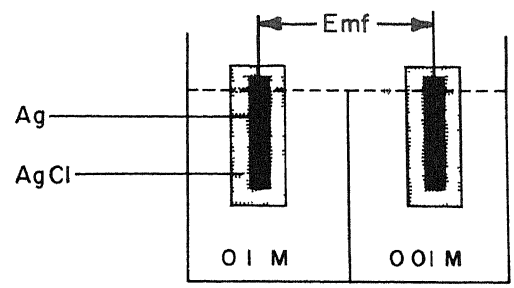


Fig 1 24 A concentration cell Both solutions contain sodium chloride The partition between the compartments is freely permeable to water and salt

diffusion of ions across the membrane or at any junction between such solutions They are therefore called *diffusion*, *membrane* or *junction* potentials The sodium chloride in the solution near the membrane tends to diffuse across into the other solution, the sodium ion diffusing faster than the chloride ion The emf at the membrane depends upon the rates (mobilities) at which the sodium and chloride ions move and the relative concentration of sodium chloride Since Na^+ diffuses faster than Cl^-, a separation of charge is produced, the solution into which a net flow of Na^+ occurs becoming positive At biological membranes, somewhat special hindrances to the movement of ions are present Potassium and chloride usually diffuse at the same rate in solution, however, the ability of chloride to move across the cell membrane is more limited than that of potassium Under these circumstances, potassium diffuses ahead of chloride and an excess of potassium ions appears on one side of the membrane Although the excess is small, it is sufficient to set up an emf or potential difference so that the solution containing the lesser concentration of potassium chloride becomes positive to that containing the higher concentration There is, in effect, a greater loss of potassium from the more concentrated to the less concentrated solution The more concentrated solution becomes negative since it has lost positive charge The concentration of K^+ increases until the emf has become sufficiently negative to keep any further K^+ from diffusing across the membrane The diffusion potential at a membrane is given by a formula similar to the Nernst equation in which the latter has been modified to take into account the mobilities of the ions The equation is

$$E = [(u - v)/(u + v)]\ 2\,303\ (RT/F) \log (C_1/C_2)$$

where the electrolyte is univalent, u and v are the mobilities of the cation and anion respectively, i e , the rates at which the ions move under unit field strength in solution The remaining symbols are the same as those in the Nernst equation Note that when the anion and cation mobilities are equal, $u = v$, and the emf is zero

Transmembrane Potential In recent years, methods have been developed for measuring the potential difference across the cell membrane at rest and during the activity associated with impulse generation and propagation In some cells, such as the giant axon of the squid it is possible to insert an electrode into the body of the axon down its length and by placing another electrode outside the cell to measure the transmembrane potential directly In other cells, where this technique is not feasible, another method is used A glass capillary is drawn out to a fine tip less than 1 μ in diameter The capillary is filled with saturated KCl or NaCl and serves as a microelectrode It is inserted directly into the soma, muscle fiber or even peripheral nerve fiber The potential difference is measured between the microelectrode and a large nonpolarizable electrode (connected via a KCl bridge) located outside the cell, this is the transmembrane potential Such a method always involves the possibility that the injury caused by the puncture will result in a gradual fall in the potential and death of the cell The membrane potentials of a variety of cells are shown in Table 1 4, they range from 61 to 94 mv with the cell interior negative in relation to the external medium

THEORY OF THE MEMBRANE POTENTIAL

The original theory of the membrane potential, largely valid today in broad outline, was first enunciated by Bernstein (1912) at the turn of this century It is well known that the concentration of potassium is much higher inside cells than outside Table 1 5 gives the value of the ratio of inside to outside concentrations for potassium and other important ions in various tissues The ratios for K^+ vary from 23 to 68 Bernstein maintained that the membrane potential was the result of the outward diffusion

Table 1 4
Resting and action potentials

Tissue	Resting Potential	Action Potential
	mv	*mv*
Loligo axon	61	96
Sepia axon	62	122
Carcinus axon	82	134
Frog myelinated nerve fiber	71	116
Frog striated muscle fiber	88	119
Frog cardiac muscle fiber	70	90
Dog cardiac muscle fiber	90	121
Kid cardiac muscle fiber	94	135

Table 1 5
Ionic content of nerve and muscle cells

Tissue	Sodium			Potassium			Chloride		
	In	Out	Ratio	In	Out	Ratio	In	Out	Ratio
	mM	*mM*		*mM*	*mM*		*mM*	*mM*	
Carcinus nerve		460		380	10	38		540	
Carcinus nerve		460		230	10	23		540	
Frog nerve (Nov)	37	120	0 31	110	2 5	44		120	
Frog nerve (Mar)				170		68			
Frog sartorius muscle	15	120	0 12	125	2 5	50	1 2	120	0 01
Frog sartorius muscle	26	120	0 22	115	2 5	46	11	120	0 092
Rat cardiac muscle	13	150	0 087	140	2 7	52		140	
Dog skeletal muscle	12	150	0 08	140	2 7	48		140	
Ionic fluxes of resting membrane									
				$\mu\mu$ *mole cm* $^{-2}$ *sec* $^{-1}$					
Sepia axon	61	31		17	58				
				(11)	(33)				
Carcinus axon				19	22				
Frog sartorius muscle	13	16		7	5				
Frog sartorius muscle		5–10			20				
Frog abdominal muscle		5		10	10				
Frog ext long dig IV muscle[a]				4	5				

[a] Extensor longus digitorum

of potassium ions from the cells Since the resting potential is a diffusion potential in which the mobility of the chloride anion was taken as zero and sodium was also presumed to be unable to penetrate the cell, the magnitude of the membrane potential can be calculated from the formula for the diffusion potential in which v 0 The formula reduces to that of Nernst,

$$E \quad 2\ 3\,(RT/F)\log\,(K_{\text{outside}}/K_{\text{inside}})$$

Partial verification for this theory has been obtained by Hodgkin and Huxley (1952b) in several experiments in which the external potassium concentration was varied and resting potential of the squid axon measured It was shown that the resting potential varied directly with the logarithm of the external concentration of potassium over a wide range Table 1 5 shows that the sodium is largely present outside cells The resting potential, V_r, was shown to be independent of the external sodium concentration, $[Na_{outside}]$, in a series of experiments on the squid axon in which $[Na_{outside}]$ was varied over a wide range with no effect on V_r Similar findings have been encountered in experiments on frog cardiac and skeletal muscle

IONIC DISTRIBUTIONS AND THE MEMBRANE POTENTIAL

If no other ion but potassium could penetrate the membrane, the membrane potential would, as noted above, be given by the equation for a diffusion potential in which the anion mobility is zero The resulting equation is then equal to the Nernst equation for a concentration cell Actually, however, potassium does not pass through the membrane alone, there is always some flux of the sodium and chloride ions Two schemes for the membrane potential have been proposed to take into account the flow through the membrane of ions in addition to potassium These are the Goldman equation and the Hodgkin-Huxley equivalent circuit for the resting membrane Tests of these equations are usually made by plotting the membrane potential against the external potassium concentration and observing how closely the equations agree with the experimentally derived curve All the equations, the Nernst relation included, are satisfactory over some range of potassium concentrations .

In order to understand the derivation and significance of the Nernst equation, let us con-

sider the problem of moving an ion in a solution in which a potential difference is present Work must be done with or against two forces, the electrical field, and any concentration difference in the ions Thus, both "electrical" and "concentration" work must be performed The electrical work (in joules) necessary to move 1 mole of a univalent ion (i e , one equivalent) against a potential difference, E (in volts), is given by the expression work = FE, where F is the Faraday (96,494 coul /mole) The concentration work required to move a mole of ion from a concentration C_1 to a higher concentration C_2 is given by the expression work = 2 303 RT log C_2/C_1 where R is the gas constant and T the absolute temperature The total work is given by the sum of the electrical and the concentration work, or work = FE + 2 303 RT log C_2/C_1 When the system is at equilibrium, the work required is zero (this is the definition of thermodynamic equilibrium) and the potential difference is given by the Nernst equation,

$$E = 2\,303\,(RT/F)\log(C_1/C_2)$$

The physical interpretation of this equation is that the tendency of an ion to diffuse down its concentration gradient is countered by the buildup of an electric field at the junction of the two solutions The direction of the field is such as to hold back the ion from further movement Note also that it is only necessary to move a very minute number of ions in order to produce the restraint required to obtain equilibrium of electric and concentration gradients As an example of an application of the Nernst equation, we calculate the transmembrane potential difference which might exist in frog muscle on the assumption that the muscle cell membrane is permeable only to K^+ The internal concentration of potassium is 155 mEq /liter, the external concentration is 4 mEq /liter Substituting these values into the Nernst equation at a temperature of 27° C, we obtain for the potential difference,

$$E = (2.303\ (8\ 2)\ (300)/96{,}500)\log(4/155)$$

$$= -95 \text{ mv}$$

The Goldman equation will not be derived here However, in the derivation, the assumptions are made that the total flow of current through the membrane is zero, that is, that the flow of negative charges is equal to the flow of positive charges, and also that the drop in potential across the membrane is linear Using these assumptions it can be shown that the resting potential will be given by the relation

$$E = 2.303\,\frac{RT}{F}\log\frac{P_K C'_K + P_{Na} C'_{Na} + P_{Cl} C_{Cl}}{P_K C_K + P_{Na} C_{Na} + P_{Cl} C'_{Cl}}$$

In this equation R, T, and F have their usual significance P represents the permeability of the membrane to the ion and the superscript '0' above the concentration i e C' represents the concentration outside the cell, the unlabeled C, the concentration within the cell When the permeabilities to Na^+ and Cl^- are taken to be zero, the equation reduces to the Nernst equation If permeabilities are assumed to have the ratio $P_K : P_{Na} : P_{Cl} = 1 : 0.01 : 0.45$ the emf of the membrane as given by the equation agrees rather well with the value measured in the squid axon, and the resting membrane potential agrees well with the calculated value over a more than 50-fold variation of concentration of external potassium

The Hodgkin-Huxley equivalent circuit constitutes a third approach to the analysis of the nerve cell membrane potential The membrane is assumed to contain separate channels through which each of the ions pass without interference from the others The total electrical current flow through the membrane is again assumed to be zero Each ion, in passing through its channel, encounters resistance to its movement through the membrane Three such channels are shown, one for Na^+, K^+, and Cl^- (Fig 1 25), each is represented by a battery (equilibrium potential) whose emf is calculated from the Nernst relation and whose opposition to current flow is shown as an electrical resistance Often the reciprocal of resistance, the conductance (g), is used to represent the channel permeability If the total current from the membrane is zero, it can be shown that the membrane potential, E, resulting from the three emfs in parallel is given by the equation

$$E = \frac{E_K g_K + E_{Na} g_{Na} + E_{Cl} g_{Cl}}{g_{Na} + g_K + g_{Cl}}$$

The g's represent the conductance of the membrane, they correspond approximately to the permeabilities to each ion. The emfs are the equilibrium potentials for each ion Thus, in frog muscle the equilibrium potential for potassium is −95 mv , for sodium +65 mv , and for chloride −95 mv , values which have been cal-

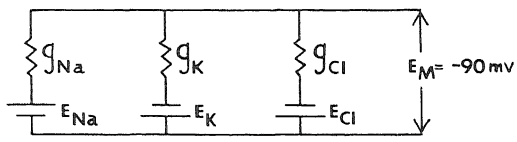


Fig 1 25 The Hodgkin-Huxley equivalent circuit for the membrane The symbols *E* represent the equilibrium potentials as determined for each ion from the Nernst equation The symbol *q* represents the conductance of each ion

culated for each ion from the Nernst equation and the concentration of ions inside and outside the cell Assuming further that the potassium conductance is 100 times greater than the Na and Cl conductances (i e , that $g_K = 100\ g_{Na} = 100\ g_{Cl}$), which is approximately true for frog muscle, we obtain for the emf across the membrane,

$$\frac{-95(g) + 65(g/100) + (-90)\,(g/100)}{g + g/100 + g/100}$$

or −95 mv , a value very close to the actual membrane potential The Hodgkin-Huxley equivalent circuit will prove very important in discussing the mechanism of the action potential in later sections

Excitability and Conductivity of the Nerve Fiber

The generation and transmission of a nerve impulse involves two conceptually independent but operationally related processes, excitation and conduction

When a nerve is stimulated, electrical events which are not propagated occur in the membrane in the vicinity of the electrodes If the local events have particular characteristics (described below), the membrane potential undergoes an abrupt change, termed the *action potential*, which is self-propagated along the axon Excitation refers to the events leading to the generation of an action potential, conduction refers to the propagation of the action potential which proceeds away from the site of excitation, much as a wave travels in a taut string It is fundamental that nerve may be excited anywhere along its length and propagation is away from the point of stimulation in both directions However, when a junction (synapse) intervenes between nerve fibers in the mammalian nervous system, then propagation can only continue in one direction Experimentally many different kinds of stimuli (electrical, thermal, mechanical, chemical) may be utilized to excite nerve or muscle, however, all of those stimuli operate by depolarizing the nerve fiber Since electrical stimuli of any intensity, shape (wave form) and duration may be easily pro duced both accurately and repetitively, this form of stimulation is used universally to study the phenomena of excitation and propagation Moreover, the nerve impulse is electrical in nature and many of its effects as an excitatory agent can be simulated by the electrical stimulus

CHARACTERISTICS OF THE STIMULUS

Nerve responds to electrical stimulation provided that the electrical stimulus fulfills certain specific criteria It must be of sufficient *intensity* and *duration* to reduce the transmembrane potential from its resting value to a critical voltage which, when achieved, results in the development of a propagated impulse This reduction in membrane potential is termed *depolarization*, and the critical voltage required for impulse propagation is termed the *threshold voltage* or *critical firing potential* The relationship between the intensity and duration of the initial stimulus is further discussed below A current just adequate to cause an impulse is called a *threshold* stimulus Intensities below threshold are referred to as *subliminal* Thresholds vary only slightly if the temperature and external ionic composition are maintained constant

Another characteristic of the stimulus is its rate of rise If the current is increased too slowly the nerve will not respond Figure 1 26 shows two linearly rising currents one of which is able to reach threshold The other current rises too slowly and the nerve is able to accommodate to the passage of the current *Accommodation*, therefore, consists of a rise in threshold of the tissue during stimulation To minimize accommodation, it is convenient to employ stimulus currents which rise extremely rapidly Two such stimuli are shown in Figure 1 27, the square wave and the exponential pulse

Even the rapidly rising pulses of Figure 1 27, if too short in duration would not result in an impulse The two properties, intensity and duration obviously interact and it is therefore

important to be aware of the relationship between the threshold stimulus intensity and the duration of the stimulus—the *strength-duration* relationship

For this purpose, the following kind of experiment is usually performed A stimulus of a fixed duration, e g , 1 msec , is applied to a nerve through two electrodes one of which is the cathode (−), the other the anode (+) The threshold stimulus is determined by increasing the current until a response is obtained—at the particular duration selected A series of other durations are selected and the thresholds determined The thresholds obtained at each duration are plotted as a function of the duration The curve so obtained is called the strength-duration relationship for nerve (Fig 1 28)

The curve is accurately described over most of its course by the empirical relationship $I = I_0 (1 - e^{-kt})^{-1}$, where I_0 and k are constants, t is the stimulus duration and I the threshold current For short durations, in which accommodation is presumably slight the relation is approximated by the equation, It = constant

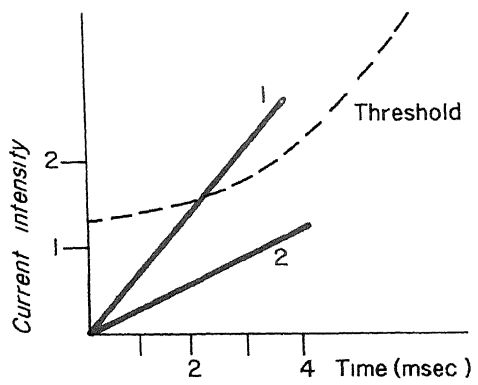


Fig 1 26 A schematic representation of the effect of rate of rise of current Curves 1 and 2 represent two stimuli with different rates of rise Stimulus 2 never reaches threshold (dashed line) Stimulus 1 attains threshold

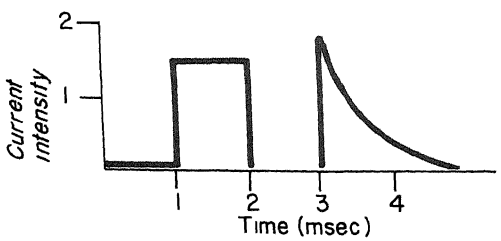


Fig 1 27 Two commonly used stimuli a rectangular pulse (*left*) and an exponential pulse (*right*)

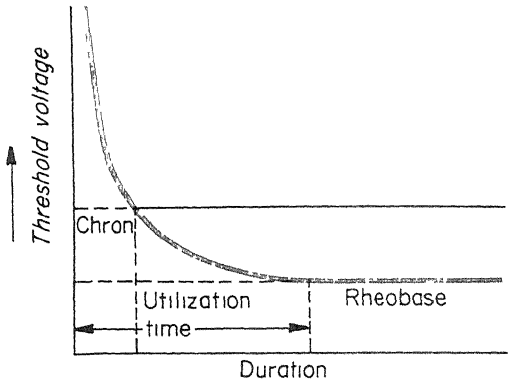


Fig 1 28 Strength duration curve *chron* chronaxie

This relationship may be interpreted as follows The current I is the charge per unit time which is placed on the membrane in a time t The product of the current and time, is therefore equivalent to a constant charge The equation implies that a critical amount of charge must be placed on the membrane, whatever the current or duration

What is the meaning of this critical amount of charge? Placing a charge on the membrane by an appropriately oriented stimulus current reduces the net charge on the membrane, i e , part of the charge on the membrane is neutralized Such a partial neutralization of charge is equivalent to a *depolarization* of the membrane, i e , to a reduction in the membrane potential from its resting negative value toward 0 Increased polarization or *hyperpolarization* results in an increase in the membrane potential, i e , increasing negativity Depolarization decreases the stability of the membrane, hyperpolarization increases its stability

A feature of excitation discovered quite early was the fact that the nerve impulse originated at the cathodal stimulating electrode, i e , at the negative stimulating electrode Figure 1 29 illustrates the lines of current that flow into and out of the electrodes At the cathode, the current flows outward through the membrane, at the anode, where current is flowing inward, excitation is hindered Continuous current flowing into the nerve depolarizes the membrane in the vicinity of the cathode and hyperpolarizes it in the vicinity of the anode Thus, it is easier to stimulate a nerve in the vicinity of the cathode because the membrane potential has been lowered, conversely, at the anode, the membrane potential is increased and higher currents are required to excite the underlying

tissue These effects of direct current on the excitability of nerve and muscle are referred to as electrotonic phenomena, the depolarization at the cathode is referred to as *catelectrotonus*, and the hyperpolarization at the anode *anelectrotonus*

SUMMATION AND THE LOCAL EXCITATORY STATE

The application of a brief subthreshold stimulus has a residual effect on nerve even though an impulse is not elicited This residual effect is revealed by the fact that the application of a second subthreshold stimulus within a millisecond can elicit a response The longer the interval between the two stimuli, the more intense the second stimulus must be to yield a response The important point, however, is the fact that two subthreshold stimuli can sum their effect on the nerve membrane so that a response is evoked The interpretation of this experiment is that the first subthreshold stimulus produces a change in the membrane potential which lasts for a millisecond or more and that this change facilitates the effect of a second stimulus The change in the membrane caused by the first stimulus is referred to as the "local excitatory state" in the vicinity of the electrode It is a nonpropagated response of the membrane which is associated with the depolarization of the membrane in the region of the cathode (see below)

The Local Excitatory State. The experiments on subthreshold summation and the electrotonic properties of nerve led to the expectation that the membrane potential should be decreased in the vicinity of the cathode, and indeed, this expectation has been affirmed by direct experiments in which the changes in membrane potential were measured in the vicinity of the cathode and anode

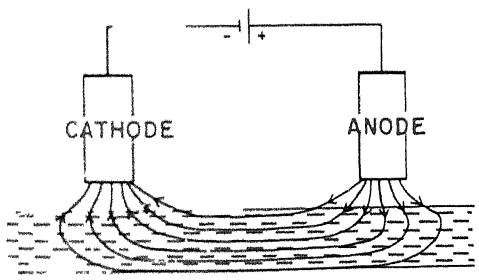


Fig 1.29 Illustrating the passage of direct current through tissue Anode is defined as the electrode which sends current *into* tissue

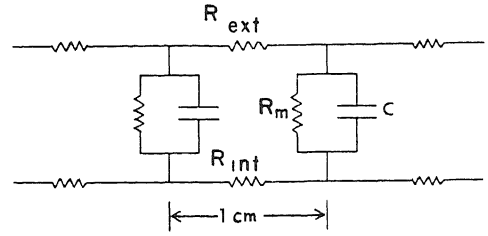


Fig 1 30 Equivalent circuit of a nerve (radius 1 μ) considered as an electric cable R_{ext} is the resistance of 1 cm of the external medium generally low R_{int} is the resistance of 1 cm of the internal axoplasm ($10^7\ \Omega$) and R_m is the resistance of 1 cm of the membrane to radial currents—of the order of $10^{10}\Omega$ The capacity C is 10^{-7} μfd for 1 cm of nerve

It is necessary first to consider the purely passive changes in the membrane potential which result from the fact that nerve has the properties of an electric cable, namely electrical resistance and capacitance (the ability to store charge) The capacitance, C, is defined as the charge, Q, which must be placed on two surfaces in order to produce a unit potential difference, V, between them ($C = Q/V$) The units of C are farads when Q is expressed in coulombs and V in volts The capacitance of the nerve membrane is 1 μf /cm ^{2}

For each unit length of nerve, there is also an external electrical resistance and an internal cytoplasmic resistance The membrane resistance constitutes a third electrical resistance Each unit length of nerve may be considered as an electrical cable (Fig 1 30) consisting of these resistances and a condenser (the capacitance) If a stimulating current is passed through a section of nerve, its condensers will charge but the most distant ones will be least charged since more external and internal resistance is included between them and the source of current If one were to measure the charging process at any moment, the voltage would be highest at the electrodes and would decrease exponentially as one proceeded away from the stimulating electrodes (Fig 1 31) This charging process is almost instantaneous and represents a nonpropagated buildup of charge along the nerve It extends for some distance along the nerve but it is never propagated as a wave

When nerve is stimulated, two electrical events take place The membrane charges passively as a cable and, at the same time, the nerve begins to react physiologically at the cathode Events at the anode are purely passive

and since excitation does not occur there, the time course of charging at the anode may be taken as that for the passive or physical charging of nerve At the cathode, passive charging occurs and, in addition, an active process which we have called the local excitatory process If the passive process at the cathode is subtracted graphically from the overall recorded response, the local excitatory process should be obtained The passive physical process at the anode and cathode are proportional to the stimulating current In order to measure the local responses, an experiment is performed in which the potential difference between electrodes placed at the anode and cathode is measured with respect to a distant electrode Figure 1 32 shows the results of the potential measurements and also the result of subtracting the passive response from the overall response to obtain the local potential which corresponds to an active, nonlinear process at the electrode

The significance of the local potential is that it demonstrates the time course of the depolarization at the cathode When the depolarization at this electrode reaches a critical value an action potential will be initiated which will propagate away from the electrode Such local responses can be found in many tissues For example, receptors must be depolarized before they give rise to an action potential, the potential representing the local excitatory state of the receptor is called the generator potential An active process also occurs at the neuromuscular end-plate which is called the end-plate potential, this depolarization initiates the propagated action potential of muscle At the synapses of neurons, a local, nonpropagated potential called the excitatory postsynaptic potential (EPSP) may be recorded which gives rise to the nerve impulse of the neuron The critical event in the excitation of all these cells is the local, nonpropagated depolarization When this depolarization reaches threshold magnitude, the explosive, propagated change in membrane potential called the nerve impulse or action potential is set off

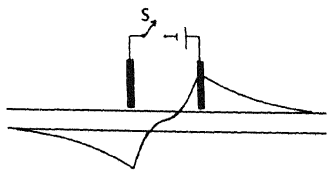


Fig 1 31 Illustrating the passive electrotonic potential along the nerve resulting from the passage of direct current into the nerve

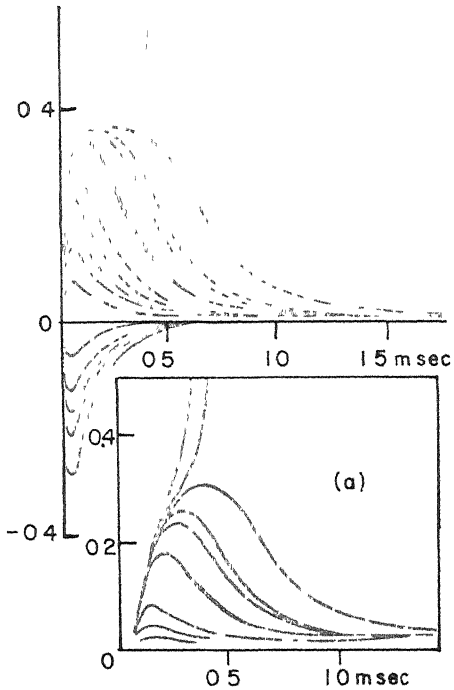


Fig 1 32 Electrical changes at stimulating electrode produced by shocks with relative strengths successively from above 1 00 (upper 6 curves), 0 96 0 85 0 71 0 57 0 43 0 21 0 21, 0 43 0 57 0 71 1 00 The ordinate scale gives the potential as a fraction of the propagated spike which was about 40 mv in amplitude The 0 96 curve is thicker than the others, because the local response has begun to fluctuate very slightly at this strength The width of the line indicates the extent of fluctuation (Inset) Responses produced by shocks with strengths, successively from above, 1 00 (upper 5 curves) 0 96 0 85 0 71 0 57 obtained from curves in upper figure by subtracting anodic changes from corresponding cathodic curves Ordinate as above (From Hodgkin, 1939)

The Nerve Impulse

RECORDING THE ACTION POTENTIAL

The external recorded action potential may be obtained by stimulating a nerve at one end and picking up the responses of the nerve some distance away with two recording electrodes and a recording device Figure 1 33 shows the arrangement by which the action potential of nerve is usually recorded with a cathode ray oscilloscope. Unlike any instrument previously employed for this purpose (e g., the string gal-

vanometer or the capillary electrometer) the moving part of the oscilloscope is a stream of electrons and consequently has virtually no mass, and thus, no inertia It is, therefore, capable of recording very rapid changes in electrical potential The instrument consists of an evacuated tube, an electron stream from a hot cathode strikes a fluorescent screen upon which it produces a spot of light On either side of the electron stream is placed a vertical plate A potential difference is created between the pair of plates, the electric field set up across the path of the stream deflects it horizontally, and sweeps it across the screen The spot of light is converted into a horizontal streak By means of a sweep oscillator the horizontal deflections are repeated many times per second A second pair of horizontal plates is placed one above, the other below, the electron stream It is this latter set of plates into which the nerve action potential is fed so that a vertical deflection of the electron stream results with production of a standing wave which is photographed and a

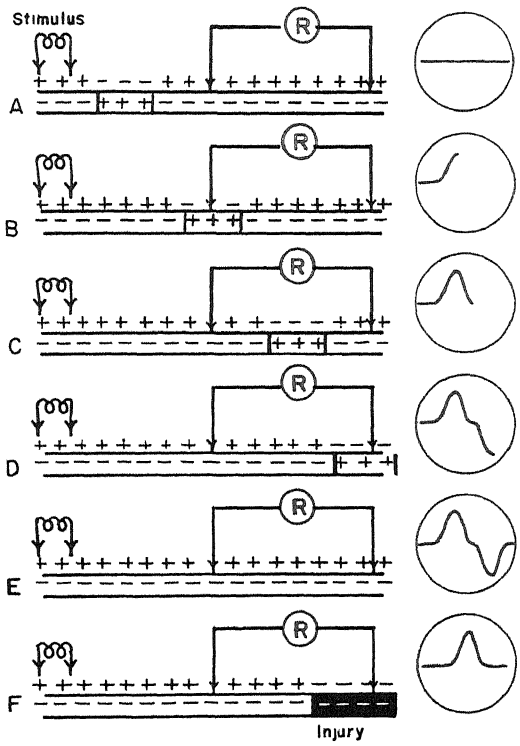


Fig 1 33 Passage of action potential down a nerve Stimulus is applied at left Cathode ray oscilloscope is indicated at *R* The recorded potential is shown at the face of the cathode ray oscilloscope at the extreme right of each figure A monophasic potential is shown in figure *F*

permanent record thus obtained The speed of the horizontal movement of the spot of light enables the time scale to be calculated, this can be varied by altering the potential applied to the horizontal pair of plates The magnitude of the action potential is determined from the height of the wave Before reaching the recording system the action current is amplified several thousand times by passing it through an amplifier This is necessary because the electron stream requires about 50 v to cause a deflection of 1 cm on the face of the tube

The response on the cathode ray oscilloscope shows up as a diphasic variation of potential The cathode ray oscilloscope records the potential under one recording electrode as a function of time with respect to the other recording electrode When the nerve is stimulated, an electrical impulse is generated which travels along the fiber The arrival of this impulse at the first electrode causes the oscilloscope beam to deflect in one direction When the impulse passes between the electrodes, no recorded potential is observed But when the impulse reaches the second electrode, a potential is recorded which is opposite in polarity to the first and accordingly, causes the oscilloscope beam to deflect in the opposite direction We, therefore, say that there are two phases present in the recorded action potential It is quite clear that one can eliminate the second phase from the recording by preventing the negativity of the impulse from reaching the second electrode This can be done by crushing the nerve either in the region between the two electrodes or directly under the second electrode Of course, if the crush is at the first electrode, no potential will be recorded at all, since the action potential cannot pass a dead region of nerve The potential recorded when the nerve is crushed at the second electrode is called a monophasic potential and is shown also in the accompanying figure (Fig 1 33F)

The Compound Nature of the Action Current Recorded from a Nerve Trunk. Erlanger and Gasser (1937) studied the action potential of mixed nerve trunks by means of the cathode ray oscillograph. They showed that the recorded "spike" is actually compounded of the individual spikes of many axons which were classified into three main types of nerve fiber—referred to as the A, B, and C groups (Fig 1 34). Several properties of nerve are correlated with the diameters of the fibers the larger the

fiber diameter, the greater the conduction velocity, the greater the magnitude of electrical response, the lower the threshold of excitation and the shorter the duration of response and the refractory period The relationship of conduction velocity to diameter of the nerve fiber is a linear one (Fig 1 35) The amplitude of the externally recorded potential is also linearly related to the fiber diameter

The A group is composed of the largest fibers, 1 to 20 μ in diameter, with conduction rates from 5 m /sec or less for the smallest fiber to 100 m /sec for the largest The fibers of the A group are all myelinated, are both sensory and motor in function, and are found in such somatic nerves as the sciatic and saphenous nerves

The B fibers are myelinated and have diameters from 1 to 3 μ, and conduction velocities from about 3 m /sec to 14 m /sec The B fibers are found solely in preganglionic autonomic nerves The C group, composed of the smallest fibers (less than 1 μ in diameter), are unmyelinated and have a conduction rate of around 2 m /sec or less, many are found in cutaneous and visceral nerves They have a high threshold, 30-fold that of the A group The A group of fibers make by far the greatest contribution to the compound action potential, and the C group the least The electrical potentials recorded from both A and C fibers exhibit slow variations in potential following the action potential, negative and positive after-potentials, but the B group do not exhibit a negative after-potential with a single response (though a negative after-potential does appear upon repetitive stimulation) The B fibers are the most susceptible to asphyxia, the C fibers, the least so

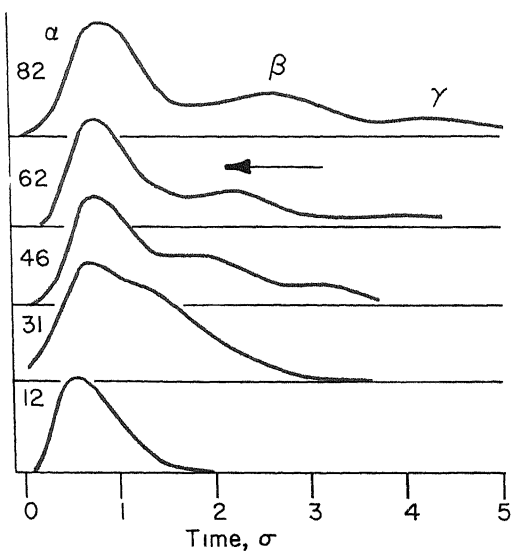


Fig 1 34 Cathode ray oscillograph records of the action currents in the sciatic nerve of the bull-frog after conduction from the point of stimulation through the distances (in millimeters) shown at the left The action potentials might be compared to runners in a race who become separated along the course as the faster contestants outstrip the slower thus in a record at 82 mm from the point of stimulation three waves are shown whereas at 12 mm the potentials are fused and only one large wave appears (Modified from Erlanger Bishop and Gasser)

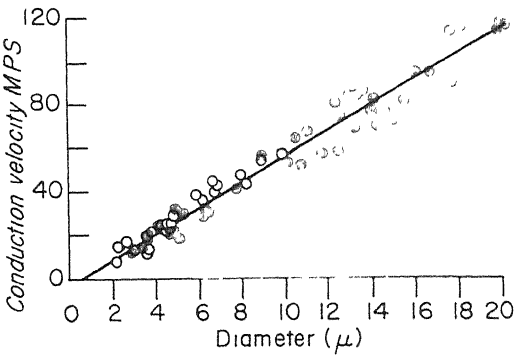


Fig 1 35 Linear relation between diameter and conduction velocity of mammalian nerve fibers Each point represents a determination of the maximum conduction velocity in meters per second and of the diameter in micra of the largest fiber of an individual nerve *Dots* adult nerves *Circles* immature nerves (After Hursh)

The linear relationship between fiber diameter and conduction velocity holds also for growing nerves of young animals During growth the diameters of the nerve fibers enlarge Conduction velocity increases proportionately so that the time taken for an impulse to travel from the toes of a kitten a few days old to the spinal cord is the same as for a full grown cat Thus, the kitten and the cat react to stimulation with about equal promptness

The diameters of the regenerating fibers in a sectioned or crushed nerve also enlarge gradually and conduction velocities increase accordingly, the relationship again being a linear one The maximum conduction velocity is not reached until maximum diameter of the fiber is attained If the axons of the nerve alone are interrupted, the sheaths of the nerve fibers remaining intact, the diameters and conduction velocities may reach those of the normal nerve, this rarely occurs if the nerve has been completely severed

CONDUCTION RATES

The velocity of the nerve impulse varies in different nerve fibers in accordance with their diameters, the thicker fibers conducting more rapidly than the fibers of smaller diameter In the large afferents from muscle spindles of the mammals, the rate is from 80 to 120 m /sec Sensory nerves of the skin being of smaller diameter have slower conduction rates Nonmedullated fibers conduct more slowly than medullated fibers Some of the fibers subserving pain sensation and those of the sympathetic nervous system have a very slow conduction rate

The following table gives the approximate conduction rates in the nerves of several different animals

Medullated nerve, mammal, 37° C 120 m /sec
Medullated nerve, dogfish, 20° C 35 m /sec
Medullated nerve, frog, 20° C 30 m /sec
Nonmedullated nerve, crab, 22° C 1 5 m /sec
Nonmedullated nerve, mammal, 37° C 1 m /sec
Nonmedullated nerve, olfactory of pike 20° C 0 2 m /sec
Nonmedullated nerve, in fishing filament of physalia, 26° C 0 12 m /sec
Nonmedullated nerve, in Anadon 0 05 m /sec
Compare the velocity of sound in air at 0° C 331 m /sec

By an indirect method of measurement the rates of conduction in various human postganglionic sympathetic nerves have been found to be from 0 85 to 2 30 m /sec The lower figures were obtained for the nerves of the leg, the higher ones for the nerves of the chest

Action Currents and Excitation. The externally recorded action potential represents a variation in electrical potential along the nerve An alternative way of looking at the nerve impulse is as a set of currents flowing out of the membrane ahead of the area of greatest depolarization or, more particularly, as the movements of these currents along the nerve (Fig 1 36) The term action current is as appropriate as action potential The advantage of understanding the currents of nerve as well as the potential becomes apparent when one considers how the self-exciting properties of the nerve impulse arise, for it is the currents themselves which act to depolarize the nerve The analogy is often made that the nerve acts like a fuse along which the ignition progresses Action currents leave the nerve ahead of the region of depolarization, acting as a virtual cathode since they have the direction of currents flowing into a cathodal electrode The hypothesis that action currents act to depolarize the region ahead of them was substantiated by experiments in which it was demonstrated that if action currents were allowed to enter but not to excite a region beyond a narcotized stretch of the nerve, then the action currents gave rise to two phenomena beyond the blocked region an increase in excitability and a depolarization of the fiber Neither the increase in excitability nor the depolarization were large enough to set up an impulse, the block diminished the intensity of the currents but it was clear from this experiment that the local currents generated by the action potential (as distinct from the action potential itself) could cause depolarization and, therefore, a change in excitability

Another hypothesis which is a part of the general theory of self-excitation holds that in order for a propagated nerve impulse to be generated, not only is a depolarization necessary, but there must also be an increase in permeability to ions at the depolarized region Thus, depolarization brings about an increased ion flow through the "not yet excited" membrane bordering on the depolarized region In other words, during the action potential a path is opened in the membrane through which the charge from resting membrane can flow Thus, by discharging resting membrane ahead of itself, the action potential becomes self-propagating. Evidence for this characteristic of self-propagating activity was obtained by Curtis and Cole (1938) who showed that there was an increased permeability to ions during activity This was demonstrated as a decrease in transverse resistance of the membrane from a resting value of 1000 Ω/cm^2 to 25 Ω/cm^2 during the rise of the action potential in the squid axon Decreases in membrane resistance have also been demonstrated in other tissues, such as mucle, during propagated activity

THE "ALL-OR-NONE" PRINCIPLE

A stimulus which is just capable of exciting a nerve fiber (threshold stimulus) sets up an impulse which is no different from one set up by a much stronger stimulus The impulse generated by the weak stimulus is conducted just as rapidly and is equal in magnitude to that generated by the strong stimulus when judged by the

action current developed or the mechanical response of the muscle it innervates Thus, the propagated disturbance established in a single nerve fiber cannot be varied by grading the intensity or duration of the stimulus, i e , the nerve fiber under a given set of conditions gives a maximal response or no response at all To make use again of the train of gunpowder analogy—the flame of a match applied to the powder fuse will start a traveling spark no less intense than one started by the flame of a torch The restoration of the strength of the impulse to its original value after passing from a narcotized region into normal nerve also shows the "all-or-none" nature of nervous conduction The well-known fact that a strong stimulus applied to a nerve *trunk* causes a compound action current of greater amplitude, and a greater muscular response than a weaker stimulus is due to the fact that the nerve trunk is composed of many fibers each of which supplies a group of muscle fibers The weak stimulus excites only a proportion of the units of the nerve trunk, whereas, a maximal stimulus excites them all For example, the cutaneous dorsi muscle of the frog is supplied by a nerve which contains only 8 or 9 fibers, each of these innervates about 20 muscle fibers When the nerve was stimulated by shocks of gradually increasing intensity, the muscular responses did not show a similar continuous rise in amplitude, on the contrary the responses of the muscle increased in a series of well-defined steps In other words, increasing the stimulus intensity produced no effect for a time upon the amplitude of the muscular response but then a slight increase in strength of stimulus produced a sudden rise in amplitude The steps which were never greater in number than the number of fibers in the nerve were due to additional fibers becoming excited as the strength of stimulus reached a certain value

It must also be remembered that the "all-or none" principle applies only for the condition of the nerve at the point where, and the moment when, the impulse arises A stimulus which will give rise to a response of a certain magnitude under one condition of the nerve

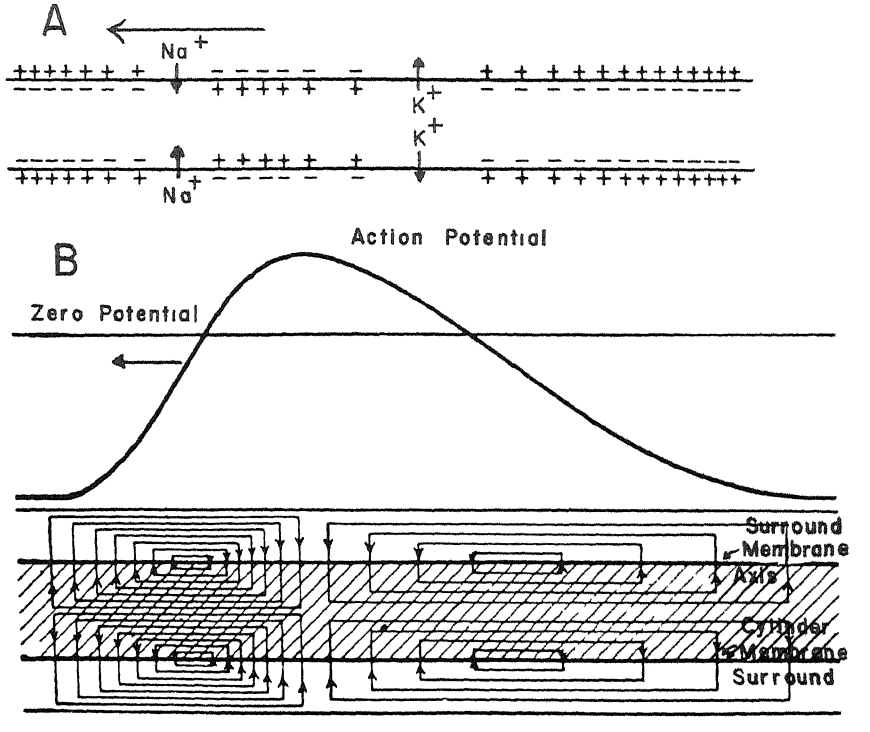


Fig 1 36 *A*) The movement of sodium and potassium during the action potential, travelling in the direction of the arrow The charge in each region is also shown *B*) Upper curve shows the potential distribution of the impulse along the nerve The lower part shows the flow of currents in the external medium and within the fiber There is a reversal of potential during the spike (After Eccles)

may give a much smaller response under other conditions, e g , during the relative refractory period (see below), narcosis, oxygen lack, etc

THE ABSOLUTE AND RELATIVE REFRACTORY PERIODS OF NERVE

For a brief interval following the passage of an impulse along the nerve fiber, a second stimulus, however strong, is unable to evoke a response This interval is called the absolute refractory period In a frog's sciatic nerve at a temperature of about 15° C the absolute refractory period has a duration of between 2 and 3 msec Its duration is roughly the same as the action potential "spike " It is much shorter in mammalian nerve (1 0 to 0 4 msec in large medullated nerve fibers)

The period during which the nerve is absolutely refractory is succeeded by one in which the nerve, though it will not respond to a stimulus of the same strength as it did before the passage of the impulse, will respond to a somewhat stronger one The excitability of the nerve gradually increases and the strength of stimulus necessary for excitation becomes progressively less (Fig 1 37) In the end, the restoration of excitability is complete and the nerve responds to a stimulus of no greater strength than that which is capable of exciting a resting nerve The period following the absolute refractory phase and during which the excitability gradually rises to normal is called the relative refractory period The time required for the excitability of the nerve to return to about 95% of its resting level ranged from 10 to 30 msec (full recovery may not be attained until the lapse of 100 msec) It should be pointed out that the failure of the nerve to conduct a second impulse is not due simply to lowered excitability at the point in the nerve where the original stimulus was applied, for during the absolute refractory period a stimulus applied to any other point upon the nerve likewise fails to set up an impulse The passage of the impulse along the nerve leaves in its wake a change of state in the membrane organization A certain time is required for the changes associated with the passage of the impulse to become reversed and the nerve restored to its resting condition

The refractory period renders a continuous excitatory state of the nerve impossible Fusion or summation of impulses does not occur The refractory period obviously must also set an

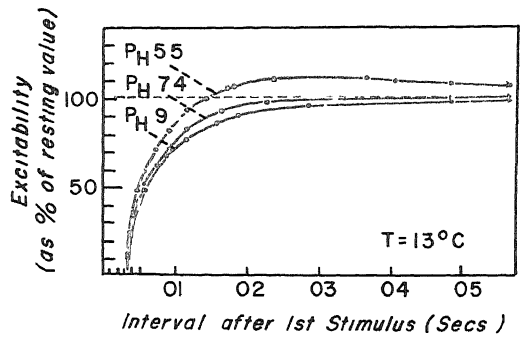


Fig 1 37 Recovery of excitability in nerve perfused with fluids of different pH (After Adrian)

upper limit on the frequency of the impulses In the mammal the absolute refractory period is about 0 5 msec The intervals between impulses cannot be shorter than the absolute refractory period, the maximum impulse frequency is around 1000/sec At this rate the impulses are travelling in the *relative* refractory period of their predecessors and are weaker and more slowly conducted In frog nerve with its refractory period of from 2 to 3 msec , the maximal impulse frequency is between 250 and 300/sec

THE CLASSICAL MECHANISM OF THE ACTION POTENTIAL

The mechanism by which the action potential is produced has not been examined but the role of the currents in exciting a region ahead of the approaching nerve impulse was noted It is now appropriate to consider the origin of the action potential in terms of the prevalent ions, namely, sodium and potassium The resting potential was shown to arise from a diffusion of potassium ions across the cell membrane The action potential has a more complex origin Recent investigations have demonstrated that the action potential results chiefly from the movement of sodium ions into the nerve fiber during the early, rising, part of the nerve impulse and that there is a movement outward of potassium ions, later during the fall of the action potential

At the turn of the century, the action potential was attributed to a complete depolarization of the membrane creating an area more negative than other parts of the nerve into which the ionic currents flowed This concept, first enunciated by Bernstein (1912), could not be tested with the methods and tissues available until 1940 At that time two groups, one in

England, the other in America, performed the critical experiments which led to a re-examination of the Bernstein concept These experiments utilized the giant axon of the squid to make a direct measurement of the transmembrane potential This axon is 0 1 mm or more in diameter so that it is possible to insert one electrode directly into it along its length and to place another electrode outside The resting potential was measured and the nerve stimulated to obtain an action potential The results of such an experiment are illustrated in Figure 1 38 According to the classical concept, the action potential arises from the fall of the membrane potential toward zero, the nerve being completely depolarized when the membrane potential reaches zero In the case of the squid axon instead of a simple depolarization of the membrane, an "overshoot" of the potential past the zero base line was noted Such an overshoot represents a reversal of the potential from the resting state so that during the peak of the action potential, the inside of the cell becomes about 50 to 60 mv positive to the outside (Fig 1 38) This is an astonishing result for, on first consideration, there does not appear to be any mechanism by which the potential difference could invert Bernstein had considered the resting potential a diffusion potential resulting from the permeability of nerve to potassium and the impermeability to sodium and chloride and the electrochemical events during the action potential were attributed to an increased flow of ions other than potassium into the nerve fibers so that the diffusion potential across the membrane disappeared when the sodium and chloride flows were permitted to take place (In a diffusion cell the emf is proportional to the difference of the mobilities of cations and anions and if these are equal, the emf is zero) This hypothesis while showing how the membrane potential could become zero did not appear to contain an explanation of how the reversal might arise

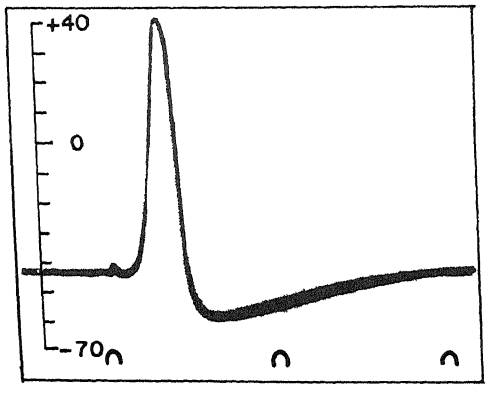


Fig 1 38 Action potential recorded with an internal electrode from a squid giant axon The scale shows the internal potential in millivolts relative to the outside bath Time marks are in 2 msec intervals (After Hodgkin and Huxley)

The Modern Concept of the Action Potential

Closer consideration of the details of ionic flow pointed to sodium as the ion which might give rise to the reversal of potential Since the concentration of sodium is higher outside than inside the nerve a flow of sodium inward would tend to make the inside positive with respect to the outside The experimental result establishing that sodium is the most important cation involved in both the production of the action potential and of the overshoot is simple The squid axon is normally surrounded by sea water osmotically equivalent to a 0 3 M sodium chloride solution If the concentration of the sodium in the sea water is changed so that less sodium is present, the amplitude of the action potential will diminish In such experiments, the resting potential is unaffected by altering the sodium concentration Accordingly it was concluded that sodium is essential to the production of the action potential and, therefore, indirectly of the overshoot Figure 1 39 illustrates the type of result obtained from such substitution experiments Note that in applying half sea water, the osmotic lack of sodium is balanced by substituting an osmotically equivalent amount of a nonelectrolyte such as sucrose or an electrolyte such as choline chloride

A variety of experiments were carried out to determine the dependence of the sodium and potassium ionic currents on the membrane potential Unfortunately, it is not possible to measure one ionic current flow during a depolarization of nerve directly and independently of another The changes in current flow affect the membrane potential which in turn affects the current flow This concept of a mutual influence of current and potential is important for the understanding of the mechanism of the action potential but it is necessary to eliminate the interaction operationally if one wishes to study the effect of a change in membrane potential on the ionic fluxes Such independence

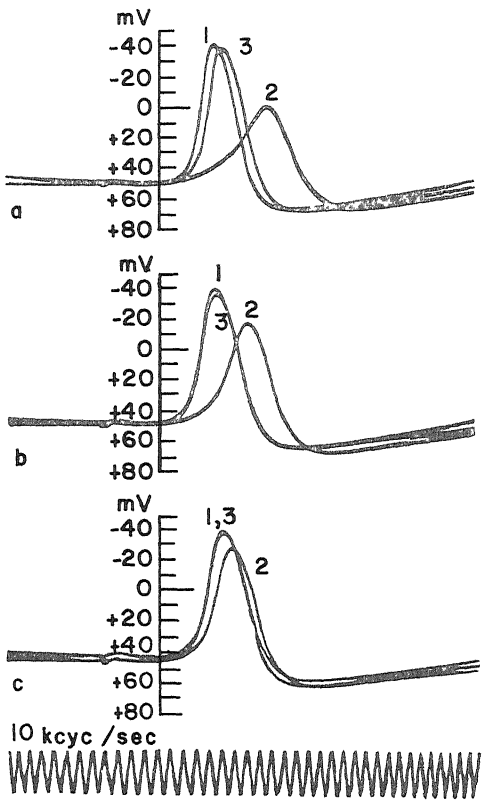


Fig 1 39 Single action potentials recorded from a squid giant axon

1) Normal action potential in sea water

2) Action potential after equilibration in medium with altered sodium *A* is 0 33 times normal *b* is 0 5 times normal and *c* is 0 71 times normal in concentration

3) Action potential after return to normal sodium solu tion

of current and potential is achieved by depolarizing the membrane to a given value and then subsequently maintaining the potential at that value The method is referred to as the *voltage clamp method* because the membrane potential is "clamped" (fixed) at an invariant value Thus, it became possible to investigate the ionic flows resulting from a given step of depolarization Note that once the potential has been clamped there will be no flow of current into the capacitive element of the membrane, since to put charge on a condenser requires a changing potential The current flow is purely resistive and ionic

In such voltage clamp experiments it has been observed that the current may change while the depolarization is being maintained This appears to violate Ohm's law, which states that the current flow in a system depends upon the voltage ($I = E/R$), and the voltage across the membrane is held constant in these experiments The explanation of these findings rests in the fact that the resistance or permeability of the membrane changes during a voltage clamp experiment, and thus, although the potential is constant, extensive changes in current occur

The effect of a small clamp on the current flow through the membrane is shown in Figure 1 40 The depolarizing clamp causes a diphasic current which is first directed inward (into the nerve fiber) and then outward If a solution which lacks sodium is substituted for the outside bath, the initial inward current disappears, but the outward current remains unaffected The early, inward current is, therefore, attributed to sodium inflow The dependence of the sodium current on the amount of depolarization is revealed by the experiments in which progressively greater amounts of depolarization are used (Fig 1 40) As the membrane potential is decreased (greater clamp voltage), the inward current increases With further increases in the depolarization, however, the sodium current decreases in magnitude and ultimately disappears

The results of a brief depolarization or clamp may be outlined as follows *1*) There occurs, initially, a brief inward current resulting from an inward movement of sodium ions, followed by an outward flow of potassium ions *2*) The amount of sodium flowing depends upon the

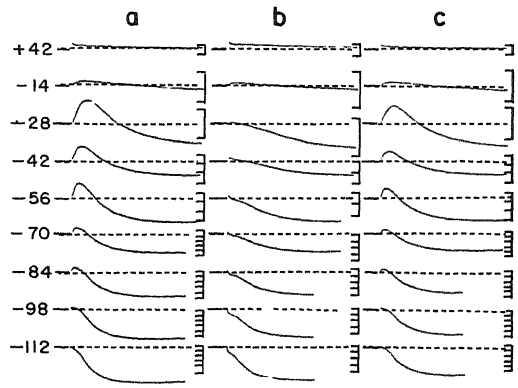


Fig 1 40 Currents flowing through squid giant axon membrane during a steady voltage clamp Figures on left are clamped values of the membrane potential in millivolts relative to resting potential (outside minus inside) Minus values represent depolarizations Columns *a* and *c* axon in sea water Column *b* axon in choline sea water without sodium Vertical scale 1 div = 0 5 ma/cm^2 Time dots are 1 msec apart (After Hodgkin and Huxley)

extent of depolarization, over a rather broad range it is seen that the greater the depolarization, the larger the sodium current .3) The potassium current is relatively unaffected by the depolarization but is greatly affected by a hyperpolarization of the membrane Increasing the membrane potential from, for example, −60 to −80 mv increases the potassium current

To understand the mechanism of the action potential, it is necessary to examine the results of the voltage clamp experiments in terms of a detailed theory—evolved by A Hodgkin (1964) and A F Huxley (1964)—which explains all the known phenomena of excitation and conduction in nerve and muscle According to this theory, the nerve membrane is represented as containing three channels through which sodium, potassium, and chloride ions may move independently The individual ions are forced through the membrane by their electrochemical potential gradients A schematic circuit representing the three channel hypothesis is shown in Figure 1 41 The value of each emf is given by the Nernst equation, the three emf's being equivalent to the three "equilibrium" potentials (as discussed above in the section on the membrane potential) It is important to understand that the network is only an electrical "equivalent" and that the channels are not physically located one next to the other Each of the channels represents a group of many identical channels in a unit area of the membrane

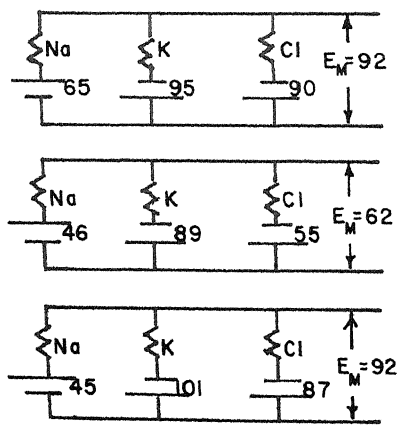


Fig 1 41 Equivalent circuits for three membranes Upper is for mammalian skeletal muscle fiber middle is for squid giant axon and lower is for frog sartorius muscle fiber E_m is resting potential all values are in millivolts

lumped together schematically as one channel The resistance signifies the opposition encountered by each ion in the unit area of membrane The reciprocal of the resistance, the conductance (g), is proportional to the permeability of the membrane to the ion

The sequence of ionic events in an action potential may now be described as follows A depolarizing voltage (stimulus) is applied to the membrane and the sodium influx increases in the direction of its concentration gradient If this Na^+ influx is greater than the flux of potassium and chloride (both of which constitute current flowing in the opposite direction), then the net sodium entry causes a change in the membrane potential But, as noted above, the change in potential across the membrane brings about an increased sodium influx, which in turn leads to a further decrease in the membrane potential Therefore, the influx of sodium builds up quite rapidly This mutual effect of sodium influx and membrane potential constitutes the *regenerative* factor in impulse transmission and is an example of a positive feedback process In terms of the equivalent circuit of Figure 1 41, the action potential develops in the following way The increase in permeability to sodium is essentially a decrease in resistance The potential difference across the membrane, which is the resultant of the three emf's in parallel, will approach that emf which has the lowest resistance in series with it The resistance to potassium and chloride does not change The membrane potential therefore tends to approach the equilibrium potential of the sodium ion which, in volts, is

$$E_{Na} = 0.058 \log (Na_{outside}/Na_{inside})$$

The equilibrium potential of the sodium ion is approximated but never quite reached by the peak of the overshoot It is possible that an increased efflux of potassium and an active process of sodium extrusion accounts for the difference between the experimentally observed peak of the overshoot and the Na equilibrium potential, i e , the theoretically expected peak.

The rapid rise in sodium current does not continue for very long and, as shown by the clamp experiments, the sodium currents are rapidly followed by a potassium efflux from the axon. The sodium influx is terminated by a process called *sodium inactivation* which develops during the increase in sodium flux Evidence for the existance of such an inactivation

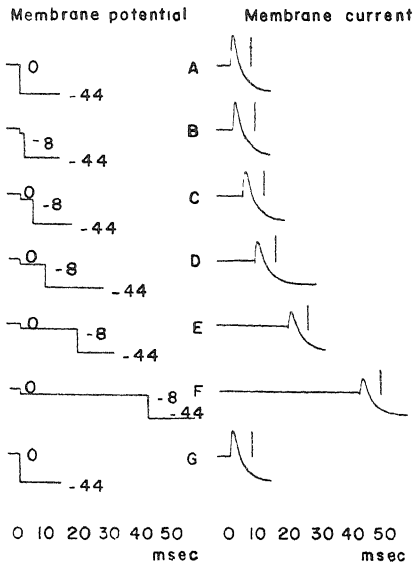


Fig 1 42 Development of inactivation during constant depolarization of 8 mv Left column displacement of membrane potential from the resting value in mv Right column membrane currents as a function of time for the displacement in the left column The vertical lines show the sodium current expected in the absence of the conditioning step (After Hodgkin and Huxley)

process has been adduced in clamp experiments of another type For example, if one applies a clamp of the order of 10 mv , this voltage change is insufficient to cause a sodium influx, yet it can nevertheless set into action the process of inactivation—as demonstrated by the fact that, following the small depolarization, a larger depolarization of about 40 mv does not evoke the magnitude of sodium current that would have developed if the smaller voltage had not been applied initially In other words the small initial voltage clamps rendered the larger subsequent voltage clamp less effective in changing the permeability of the membrane to Na^+ Thus, by following small clamps by larger clamps at various intervals, the time course of development of the inactivation can be ascertained (Fig 1 42) Thus, the inactivation process turns off the sodium current and, thereby contributes toward restoration of the resting state

A second process also aids in the restoration of the resting membrane potential It was noted that a potassium efflux follows the sodium influx The time course of efflux of potassium ions corresponds to the fall of the action potential The increase in conductance to potassium means that the change in potential across the membrane is now being driven by the electrochemical potential due to the potassium gradient This net efflux of cations, the consequence of the increased permeability to potassium ions (sometimes referred to as rectification) tends to restore the membrane potential to its original state of internal negativity Figure 1 43 shows the time course of the increase in conductance of potassium in relation to the sodium inactivation process

Both the sodium and potassium flows during the phase of increased permeability are in the direction of their concentration gradients The flow of ions during the action potential is purely passive and depends upon the concentration gradient The production of a nerve impulse should therefore be independent of metabolism This deduction is substantiated by experiments in which both aerobic and anaerobic processes are blocked by inhibitors Nerves may continue to conduct for hours even when stimulated at high rates under the influence of metabolic inhibitors The increased permeability to sodium and potassium persists although the means for pumping sodium and potassium have been cut off This is not to say that transport processes are not important for maintaining the ionic inequalities If nerve is continually stimulated while being treated with an inhibitor, it fails to pump out the sodium which accumulates inside the nerve both during the resting state and during excitation—and it also fails to reaccumulate potassium to restore its resting intracellular K^+ concentration Ultimately, therefore, the nerve will depolarize as the result of potassium loss and/or sodium influx Thus, the ion pumps and associated metabolic activity which maintains the steady state ionic gradients across the nerve cell membrane

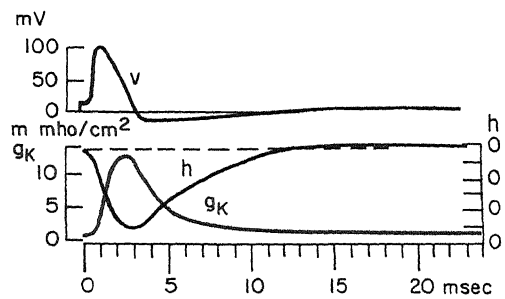


Fig 1 43 Time course of inactivation (h) and potassium conductance (g_K) during a nonpropagated action potential Upper curve is action potential in response to a 15 mv initial depolarization (After Hodgkin and Huxley)

are necessary in the long run, but neither the resting potential nor the action potential depend primarily upon metabolism for their production

How does this theory of the action potential explain other important phenomena of nerve such as refractoriness and accommodation? The absolute refractory period is the result of the inability of nerve to respond to a stimulus no matter how intense This inability corresponds to the period during the action potential when sodium influx is completely inactivated It is impossible to turn on the sodium influx machinery in any degree whatever at the height of the inactivation process During the relative refractory period, the nerve is under the influence of partial sodium inactivation and increased potassium efflux The potassium efflux tends to maintain the membrane potential at the original unexcited level The restorative processes of sodium inactivation and potassium efflux (rectification) must be opposed by very large depolarizations in order to get sufficient sodium influx to start regeneration Similarly, accommodation, which is a rise in threshold during the persistent application of a linearly rising current, may be explained as the result of growth of inactivation at the cathode of an exciting current, much as inactivation developed in the clamp experiments which demonstrated the existence of the process

Little is known concerning the molecular events within the nerve membrane that are triggered by depolarization and then lead to increased sodium conductance It is clear, however, that these events are reversible A hypothesis, advanced by Tasaki (1968), holds that the subunit organization of the membrane may exist in one of two stable states, "resting" or "excited," and that calcium ions within the membrane are critical in regulating a reversible transition between those states In this view, depolarization allows monovalent cations to enter the membrane displacing the divalent calcium ions from their anionic binding sites This displacement of calcium ions permits a transition of the membrane to a state of increased permeability, possibly through alterations in peptide cross-linkages This concept is supported by experiments with barnacle muscle and perfused squid axon in which it has been demonstrated that calcium ions are essential for membrane excitability regardless of which monovalent cations are available to carry current It has also been shown that excitability is preserved in these tissues when sodium ions are perfused internally provided that calcium is present

Nerve Metabolism

INTERMEDIARY METABOLISM

Nerve, like other tissues contains the enzymatic apparatus for glycolysis the citric acid (Krebs) cycle and the electron transport system and, thus, can generate and store energy in the form of adenosine triphosphate (ATP) The pathway of glycolysis is concerned with the breakdown of glucose to pyruvate and/or lactate Only a small amount of ATP is produced in this pathway which is anaerobic The major part of the ATP is produced aerobically in the Krebs cycle and the electron transport system (for details see discussion of the energy generating systems of cells in chap 1

Under anaerobic conditions, e g , in an atmosphere of nitrogen, resting nerve converts glucose to lactate quantitatively However, the nerve at rest produces very little lactic acid aerobically, although recent experiments have shown a clear cut increase in lactic acid production in an oxygen atmosphere (aerobic glycolysis) when the nerve is stimulated at 100/sec

Nerve conduction continues under anaerobic conditions (nitrogen atmosphere), for, as noted previously, the action potential does not depend immediately upon metabolism Ultimately, of course, conduction will fail under anaerobic conditions because of the gradual accumulation of sodium and consequent depolarization of the nerve fiber

Further evidence for the existence of glycolytic systems in nerve comes from the effect of metabolic inhibitors on nerve metabolism Iodoacetic acid, which interferes with glycolysis, causes a decrease in the oxygen consumption of nerve At the same time the effect of the inhibitor may be slowed by the administration of sodium lactate, the latter probably forming pyruvate which is then oxidized The cofactor necessary for the formation of lactic acid from pyruvic acid, diphosphopyridine nucleotide (DPN or nicotinamide adenine dinucleotide, NAD), is found in frog nerve It has also been shown that the administration of glucose partially prevents the loss of potassium and gain of sodium resulting from anoxia of nerve

Evidence for the presence of the Krebs cycle

in nerve comes from the effect of inhibitors on the oxygen consumption of frog nerve Methylfluoracetate, an inhibitor which acts in the Krebs cycle, inhibits the rate of oxygen uptake and ultimately causes a failure of conduction If the nerve is pre-equilibrated with sodium fumarate or sodium succinate, the inhibition is overcome and the nerve regains its ability to conduct and to consume oxygen Fumarate and succinate are intermediates in the Krebs cycle and serve to keep going the portion of the cycle devoted to converting succinate or fumarate to carbon dioxide and water (see energy generating systems, chap 5)

The importance of the electron transport scheme has been demonstrated by experiments in which the cytochrome oxidase of nerve has been inactivated by carbon monoxide When nerve is poisoned with carbon monoxide, the cytochrome oxidase fails to function and molecular oxygen cannot be reduced for its ultimate reaction with metabolically-derived hydrogen to form water The nerve behaves as though it were in nitrogen However, it is possible to dissociate the carbon monoxide from the cytochrome oxidase by shining visible light on the nerve The nerve then regains its ability to consume oxygen and will conduct an action potential

RESTING AND ACTIVE METABOLISM

In the resting state, the oxygen consumption of frog sciatic nerve is quite low, 30 to 40 mm ^{3}/g wet weight/hr in contrast to that of mammalian nerve which ranges from 200 to 300 mm ^{3}/g /hr These values may be compared with those of brain which is of the order of 2000 mm ^{3}/g /hr Corresponding to the consumption of oxygen, there is a resting release of heat by nerve which amounts to 0 15 cal /g /hr for frog sciatic nerve

The oxygen consumption of nerve in vitro is accompanied by a release of carbon dioxide to the extent that the respiratory quotient (moles CO_2/moles O_2) is 0 8 This value indicates that the fuel of nerve may not be exclusively glucose which would yield a respiratory quotient R Q of 1 as indeed proves to be the case when the R Q. of mammalian brain is measured in vivo During excitation the respiratory quotient of the extra oxygen consumption of nerve changes to 0 9, suggesting that the recovery processes following excitation use a different substrate than is used during the resting state

How much of the resting heat production of nerve can be attributed to the heat production of the process pumping Na^+ out of nerve and K^+ into the cell during the resting state? One can calculate the energy required to extrude sodium against both the electrical forces in the membrane and the concentration gradient A similar calculation may be made for potassium uptake If the process is assumed to be inefficient and to produce heat equal to the work done (50% mechanical efficiency) an estimate of the heat production can be obtained In the case of sodium, it is necessary to pump the ion against both the electric field and its concentration gradient since the ion is positive and it must be forced out against the negative attraction of the interior The electric work required is given by EF (ionic flux) The concentration work is obtained from the relationship $RT \log C_1/C_2$, where the symbols have their usual significance Substituting the values for frog nerve of $E = 0\ 070$ v , $F = 23{,}050$ cal /v /mole, $C_1/C_2 = 3$ (sodium) and a sodium flux of 10^{-5} mole/g /hr , one obtains an electrical work of 0 015 cal /g /hr and a concentration work of 0 007 cal /g /hr A similar calculation for potassium shows that the energy necessary to pump this ion into nerve is 0 019 cal /g /hr The work necessary to pump both sodium and potassium is 0 041 cal /g /hr If the pump has a mechanical efficiency of 50% then its operation—complete with inefficiency—accounts for 0 082 cal or about 50% of the total energy consumption of 0 15 cal /g /hr noted above The value of 0 041 cal /g /hr corresponds to 4 1 kcal /mole of Na^+ and K^+ exchanged Thus, taking 8 kcal / mole as the free energy of ATP hydrolysis, it is seen that approximately 2 moles of Na^+ can be exchanged for 2 moles of K^+ for each mole of ATP converted to adenosine diphosphate (ADP) plus P_i

When nerve is stimulated, there is an increase in the oxygen consumption and an increased loss of heat The extra oxygen consumption may double The increase in heat production and oxygen consumption parallel one another and depend upon the frequency of stimulation The increase in oxygen consumption rises gradually with increase in frequency of stimulation, leveling off at about 100 impulses/sec The extra heat which accompanies short tetanic stimulation has been shown by Hill (1938, 1959) and co-workers to consist of two phases, an early one lasting only 2 to 3

sec , called the initial heat which amounts to a few percent of the total extra heat followed by the remainder of the heat production which is called the delayed heat The latter is presumably concerned with the recovery process in nerve involving the pumping of ions The extra oxygen consumption of nerve may be eliminated by inhibitors (azide and methylfluoracetate) which have no effect on the action potential

Abbott, Hill, and Howarth (1958), using improved thermal methods have been able to measure the heat output resulting from a single shock to a crab nerve They found that the initial heat accompanying an action potential occurred in two phases, an early, rapid portion, 1 msec in duration, which was positive (exothermic), and a delayed, negative (endothermic) heat of absorption The actual heats during an impulse were $+14 \times 10^{-16}$ and -12×10^{-16} cal /g /impulse The difference gives the net initial heat production Calculations by Hodgkin indicate that the initial heat can be accounted for by the heat production of the action currents flowing within and without the nerve But other processes such as heat produced by the mixing of sodium entering nerve may also account for the initial heat

Thus, active cation transport may be the major energy-consuming process in specialized tissues such as nerve McIlwain (1966) has shown that electrical pulsation of brain slices in vitro results in cation-dependent respiratory stimulation accompanied by acceleration of active cation fluxes About 30% of the increased respiration may be accounted for by the increased cation flux. The maintenance or restoration of cation gradients in nerve as well as other tissues has been correlated with a sodium-and-potassium-stimulated adenosine triphosphatase present in membranes This enzyme is believed to utilize the chemical bond energy in ATP to translocate Na^+ and K^+ across the membrane against their respective concentration gradients

Synaptic Transmission

In the nervous system of mammals, the formation of conducting pathways does not involve direct anatomical continuity between one neuron and another or between a neuron and a cell of an effector organ An axon gives rise to many expanded terminal branches (presynaptic terminal boutons) The postsynaptic component of the synapse may be formed by any part of the surface of the second neuron with the exception generally of the axon hillocks it usually is a dendrite (axodendritic synapse) but it may be part of the cell body (axosomatic synapse) or part of the membrane of another axon (axoaxonic synapse) A single neuron may be involved in many thousands of synaptic connections but in every case the impulse transmission can occur only in one direction In the mammalian central nervous system there is a cleft of about 20 nm (the *synaptic cleft*) separating the presynaptic axon terminal (the *presynaptic membrane*) and the surface of the postsynaptic component (the *postsynaptic membrane*)

At the sites of opposition, both the presynaptic and postsynaptic membranes exhibit a plaque of electron dense material on their cytoplasmic surfaces Beneath this plaque on the presynaptic membrane and within the presynaptic bouton, there are vesicles ranging in diameter from 10 to 50 nm and often numerous mitochondria

The action potential does not cross the synaptic cleft but instead causes the release of a transmitter substance which is stored in the vesicles of the presynaptic terminals The transmitter substance then crosses the gap, attaches to "*receptors*" on the postsynaptic membrane which is then either *excited* or *inhibited* depending on the precise nature of the permeability changes ensuing after combination of the transmitter and the postsynaptic receptor system (see below)

Chemical transmission was first demonstrated by Otto Loewi in 1920 in a classical experiment in which the heart of a frog was slowed by perfusing it with blood taken from the heart of a second frog that had been slowed as a result of vagal stimulation Loewi (1921) concluded that the vagal stimulation had caused the release of a substance identified as acetylcholine—which altered the excitability of cardiac muscle

THE EXCITATORY POSTSYNAPTIC POTENTIAL (EPSP)

Eccles (1953) and co-workers have inserted finely drawn-out microelectrodes made of glass and filled with KCl or NaCl into the cell body of motoneurons of the spinal cord When the microelectrode is inserted, a resting potential of about 70 mv. is observed, with the interior of the cell negative to the outside If the moto-

neuron is antidromically excited, i e , by stimulating the axon of the motoneuron, the motoneuron is excited without the intervention of the synapse An action potential with an overshoot of about +20 mv appears Thus, antidromic excitation gives rise to an action potential in the postsynaptic cell, the motoneuron, which is no different in its general characteristics than that obtained from squid axon or the muscle fiber The slight inflections which appear on the rise do not enter into the present discussion

The events of interest in synaptic excitation, however, occur when excitation is delivered orthodromically, i e , to the afferent neuron of the dorsal root, so that the presynaptic endings of this neuron are excited In such experiments in which a synapse is present, an additional potential appears in recording from the postsynaptic cell which is a sign of postsynaptic depolarization Figure 1 44 shows the response of the motoneuron to a presynaptic volley The slow potential change which appears just before the action potential is called the excitatory postsynaptic potential or EPSP It consists of a hypopolarization of the membrane which may last for 20 msec and can best be seen when the extent of hypopolarization is insufficient to give rise to an action potential When the membrane potential falls to a critical value of about −60 mv , the motoneuron fires and an action potential appears superimposed on the EPSP Temporal summation of such EPSP's is also possible If two subliminal volleys are sent in over the same nerve, each volley produces an effect which is manifested by an EPSP The EPSP's will then sum and if the critical level of depolarization is reached, an impulse will be set off

The EPSP is monophasic and nonpropagating, it represents a hypopolarization which is localized to the soma of the motoneuron Unlike the action potential, the EPSP is a potential which is not all-or-none in character since it can be augmented simply by increasing the intensity of the input volley Moreover, the EPSP's of different inputs can sum on a postsynaptic cell to produce a greater hypopolarization These characteristics show that the EPSP is produced in a process which is fundamentally different from that of the action potential

Several lines of evidence indicate that the EPSP arises from an influx of all ions through

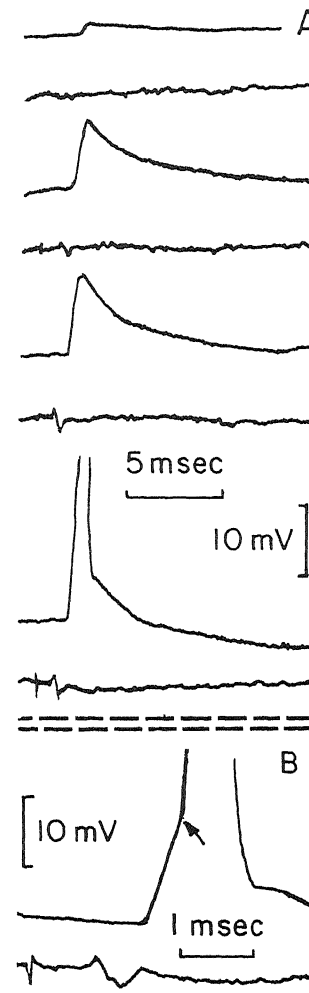


Fig 1 44 Intracellular potentials set up in a biceps-semitendinous neuron by various sizes of volleys in the afferent nerve (lower records) *A* shows synaptic potentials (upper records) of graded size the largest setting up an action potential In *B* a faster record of this response is shown Note spike arising at arrow from more slowly rising synaptic potential (Brock Coombs and Eccles 1952)

the postsynaptic membrane into the cell One group of experiments concerns setting the membrane at a given membrane potential by passing current across the membrane A double-barreled microelectrode was inserted into a motoneuron One electrode was used to hypopolarize or hyperpolarize the neuron with a direct current, the other to record the EPSP When the membrane potential was increased, the amplitude and rate of rise of the EPSP increased Conversely, when the membrane potential was decreased, the EPSP decreased

Reversal of the direction of the EPSP occurred when the membrane potential was set up at zero volts, which is therefore the equilibrium value for the EPSP The explanation of this effect goes back to the experiments on the squid axon It will be recalled that when the axon was clamped at the equilibrium potential of an ion, a reversal of current flow occurred through the membrane above and below this value of clamp voltage We may imagine that the motoneuron membrane contains channels through which ions can flow to cause the development of an EPSP These channels are supposed to be additional to those serving as producers of the action potential A diagram of the equivalent circuit for such channels is shown in Figure 1 45 When the membrane potential is set at zero, the equilibrium potential of the EPSP, a change in resistance of this channel can have no effect on the recorded potential No EPSP will develop Under normal circumstances the membrane potential is determined by the equilibrium potential for potassium but when the EPSP channels decrease in resistance, the potential recorded tends to approach the equilibrium potential for the EPSP Now the only process which can have an equilibrium potential of zero is one in which the membrane is freely permeable to all ions This process will have an equilibrium potential of zero because it is a diffusion potential and the value of the emf developed in a diffusion potential depends directly upon the difference of mobilities of the cations and anions During the production of the EPSP it is presumed that the mobilities are equal so that the process has an equilibrium potential of zero

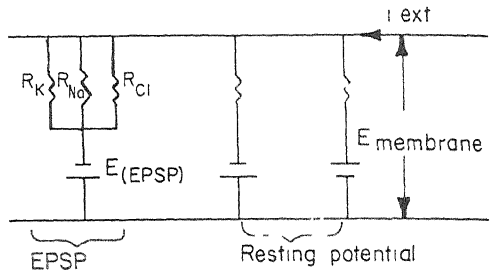


Fig 1 45 Equivalent circuit of motoneuron membrane for EPSP Left circuit represent channels giving rise to EPSP Right circuits represent circuit of polarized membrane

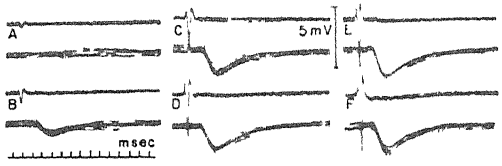


Fig 1 46 Lower records give intracellular responses of a biceps semitendinosus motoneuron to a quadriceps volley of progressively increasing size as is shown by the upper records which are recorded from the L6 dorsal root by a surface electrode (denatured deflection signalling negativity) Note three gradations in the size of the IPSP from A to B from B to C and from D to E Voltage scale gives 5 mv for intracellular records downward deflection indicating membrane hyperpolarization (from Coombs, Eccles and Fatt 1955)

THE INHIBITORY POTENTIALS

If a monosynaptic reflex is evoked by orthodromic excitation, an inhibition of the reflex can be produced by stimulating the afferent nerve from an ipsilateral antagonistic muscle as described more fully in the section on inhibition The effect of the inhibitory stimulus is to cause a hyperpolarization of the motoneuronal postsynaptic membrane in a direction opposite to that of the EPSP (Fig 1 46) This hyperpolarization consequent upon an inhibitory stimulus is called the inhibitory postsynaptic potential or IPSP Its time course is the same as that of EPSP, of the order of 20 msec , and corresponds to the curve of inhibition obtained from reflex studies Indeed the IPSP is a mirror image of the EPSP in the same motoneuron The IPSP being a hyperpolarization of the membrane renders it less excitable Any EPSP which occurs during an IPSP will generate currents which will be less effective in causing excitation of the cell

The problem of determining the nature of the currents responsible for the IPSP has been attacked by the same procedures described for the EPSP Using doubled-barrelled microelectrodes, it is found that setting the membrane potential at −80 mv will cause the disappearance of the IPSP The potential 80 mv. is the average of that determined by potassium (70) and chloride (−90) suggesting that the IPSP results from an increased flow of K^+ and Cl^- making the inside more negative. Further evidence however, points to chloride as the principal ion involved in producing the IPSP The method of electrophoretic injection has been used to demonstrate this One barrel of a double microelectrode is filled with KCl and made negative so that this electrode effectively drives chloride ions into the cell, it is a hyper-

polarizing current since it is internally negative The other barrel electrode records the IPSP It is found that the injection of chloride into the cell converts the IPSP elicited by a group I_a afferent volley from a hyperpolarization to a hypopolarization (EPSP) The injection of chloride caused an increased internal concentration of chloride The transmitter causing inhibition results in an increased permeability to chloride, but the chloride ionic gradient is now opposite to its normal direction so that a hypopolarization will occur instead of a hyperpolarization if the membrane is especially permeable to chloride Chloride now flows out and makes the inside of the cell less negative A number of anions were injected internally and all those with less than a certain hydrated ion diameter caused a reversal of the IPSP to an EPSP Apparently the inhibitory substance opens pores which allow ions below a certain size to pass Since chloride is the only anion which under in vivo conditions exists in a high enough concentration to flow in during such an inhibition, it is presumed that chloride is the main contributor to the inhibitory currents

Types of Inhibition

Some restraint must be placed upon the ability of muscle and neural circuits to respond In reaching for an object, for example, the muscle must be controlled if the movement is to be accurate and the muscle is not to overshoot the mark Certainly reciprocal innervation of some type must often be employed so that when a muscle is activated its antagonist will be inhibited, otherwise persistent opposition of an undesired sort will be encountered These two reasons give some indication of the usefulness of inhibition in motor movement A third less obvious necessity for inhibition arises from the very complexity of neural activity in which some form of "negative feedback" or inhibition is necessary in order to keep the complex neural networks from over-activity An abnormal form of such activity is observed, for example, in strychnine poisoning in which inhibitory neurons have been shown to be inactivated In a strychninized animal any stimulus leads to persistent neural activity and convulsion Again, in the disorder known as Parkinsonism, periodic activity in the form of a tremor of a limb manifests itself, presumably as a result of injury to inhibitory systems which restrain normal motor activity

Three kinds of inhibition have been extensively studied *1)* direct, *2)* presynaptic, and *3)* Renshaw cell or recurrent inhibition A fourth type of inhibition, called indirect inhibition and extensively discussed by Sherrington is not now considered to be inhibition per se but a form of occlusion

DIRECT INHIBITION

The direct form of inhibition can be observed by evoking a monosynaptic reflex and then depressing the amplitude of the reflex by stimulating an ipsilateral skin nerve (type II fibers) It is necessary to stimulate the inhibitory nerve shortly before eliciting the reflex By varying the interval between the inhibitory and excitatory stimulus, the time course of direct inhibition can be obtained The inhibitory nerve is stimulated first A curve of the form of Figure 1 47 is obtained in which maximum inhibition is obtained at an interval of 0 5 msec between the two stimuli and with a delay lasting for about 10 msec The interval 0 5 msec , is equivalent to one synaptic delay so that at least one inhibitory interneuron is interposed between the afferent neuron and the motoneuron It is believed at present that this interneuron in the inhibitory pathway possesses the ability to secrete a substance at its terminals which hyperpolarizes the membrane and leads to an IPSP Thus, a neuron which is normally excitatory to other neurons may exert an inhibitory action by the interposition of an inhibitory neuron

Direct inhibition is very susceptible to strych-

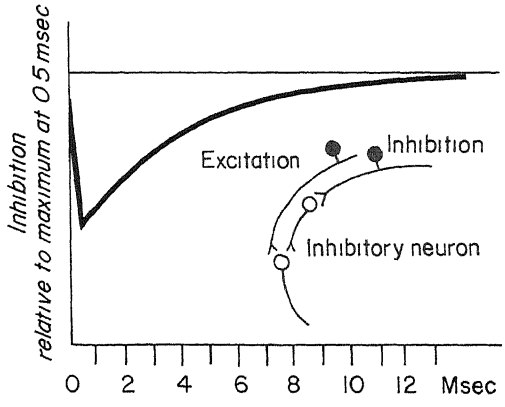


Fig 1 47 Time course of direct inhibition Insert shows circuit containing inhibitory neuron (after Lloyd)

nine injection which can completely abolish it This effect of strychnine is the basis for the explanation of its convulsant activity—all inhibition of the direct type having been removed, the slightest stimulus causes a tremendous extensor response

PRESYNAPTIC INHIBITION

Early in the studies of the electrical responses evoked by reflex action electrodes were placed on the surface of the spinal cord or on dorsal roots, and large potentials which were positive to distant regions, called P waves, were recorded It was suspected that this positivity might be the sign of some inhibitory presynaptic activity going on in the cord To demonstrate presynaptic inhibition, a reflex response is elicited in extensor motoneurons This reflex will be inhibited if it is preceded by stimuli from any nerve entering the cord Group I_a and I_b fibers are most effective in causing the inhibition but groups II and III will also serve The time course of such an inhibition is shown in Figure 1 48 Unlike direct inhibition, maximum inhibition is observed when the inhibitory stimulus precedes the excitatory by about 20 msec The inhibition may endure for as long as 200 to 300 msec and is extremely resistant to strychnine Several lines of evidence indicate that this inhibition is presynaptic Recording within the motoneuron at the time that inhibition is produced shows no changes in the membrane potential Neither a hyperpolarization or a depolarization is produced by an inhibitory volley Instead, it is noted that the magnitude of the EPSP which is set up by the reflex volley is diminished by the inhibitory stimulus and that the magnitude of the EPSP parallels the curve of inhibition (Fig 1 48) It has also been shown that there is a depolarization of the presynaptic fine terminals entering the dorsal region of the cord Microelectrodes are inserted into and just outside a nerve fiber The undesirable potential of neighboring neurons which is recorded inside a nerve fiber together with the membrane potential, may be obviated by subtracting the externally recorded potential from the internally recorded to obtain the true membrane potential It is found that the time course of the depolarization produced by inhibitory volleys parallels the time course of inhibition

Although the mechanism is not clear, depolarization in axo-axonal synapses on presyn

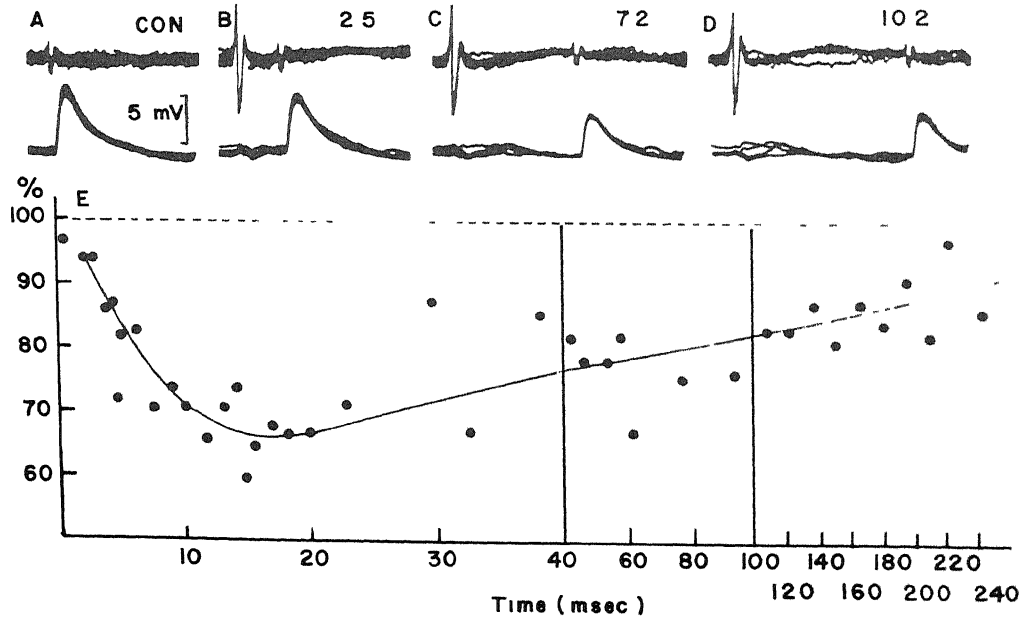


Fig 1 48 Time course of presynaptic inhibition (*E*) Inhibitory stimulus is maximal group I volley in biceps semitendinous nerve Monosynaptic EPSP s were evoked at various intervals after the inhibitory stimulus by stimulating the gastrocnemius-soleus nerve maximally *A* shows control EPSP *B C* and *D* are the EPSP's at intervals marked above the records

aptic terminals may reduce the amount of excitatory transmitter released from the presynaptic terminals, thus decreasing the EPSP amplitude Since long delays are involved in presynaptic inhibition, many interneurons are interposed between the first and last neurons of the chain

RENSHAW CELL INHIBITION

In 1946, Renshaw discovered that a volley of impulses delivered to motor axons causes an inhibition of all types of motoneurons at the segmental level This type of inhibition has therefore been called antidromic because the inhibition may be evoked by firing backwards over the motor roots into the spinal cord When such antidromic excitation was used, an afterdischarge of quite high frequency was observed in microelectrode recordings from the ventral *horn* It was also shown that the discharge did not occur in the motoneuron but in neighboring cells near the motoneuron—in cells which discharge with a high frequency when the antidromic excitation occurs At the same time the motoneuron is inhibited These neighboring cells are therefore believed to cause the inhibition The motoneuron displays a hyperpolarization which has all the characteristics of an inhibitory postsynaptic potential and the IPSP lasts for the period of time corresponding to the discharge in the neighboring cells which are called Renshaw cells An anatomical pathway has been suggested to explain these results (Fig 1 49) The motoneuron gives off a collateral to the Renshaw cells, the Renshaw cell axon returns to the motoneuron and inhibits it

The repetitive discharge of the Renshaw cell is presumed to be caused by an accumulation of acetylcholine at the junction between the motoneuron collateral fiber and the Renshaw cell It has been shown, for example, that the discharge of the Renshaw cell may be prolonged by anticholinesterases which allow acetylcholine to accumulate It is also possible to inhibit the discharge with β-dihydroerythroidine, a substance which blocks the action of acetylcholine The functional significance of this pathway is not clear It appears to serve as a general synaptic inhibitor and may act to limit the frequencies of impulses going to the motor endplate

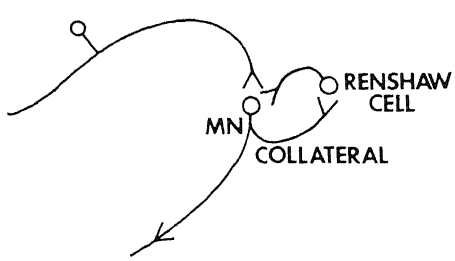


Fig 1 49 Schematic drawing of recurrent inhibition

Electrical and Chemical Transmission

The events at a synapse may be interpreted according to either a chemical or electrical theory of synaptic transmission The chemical theory of transmission attributes the response of the postsynaptic cell to a chemical substance (transmitter) released from the presynaptic nerve terminals The electrical mechanism of transmission presupposes the activation of postsynaptic neurons by electrical currents flowing out from the presynaptic terminals into the postsynaptic cell Examples of each of these types of transmission may be found in various kinds of synapses among vertebrates and invertebrates It may be stated, however, that no firm evidence of electrical transmission at mammalian synapses has as yet been obtained

The present day understanding of chemical transmission is based primarily on studies of the neuromuscular junction because of its accessibility to investigation The chemical mechanism of transmitter release and action may be schematically outlined as in Figure 1 50

The action potential, upon invasion of the presynaptic terminal, causes release of transmitter material In the case of the neuromuscular junction in which the transmitter is acetylcholine, this substance is stored in clear vesicles in the presynaptic nerve terminal In their classical experiments with frog muscle fibers, Fatt and Katz (1952) observed that under certain conditions the muscle end-plate (postsynaptic membrane) exhibited intermittent, spontaneous miniature potentials in a random manner These miniature end-plate potentials (MEPP's), of insufficient amplitude to depolarize the membrane to threshold, were related to random release of acetylcholine packets from the presynaptic vesicles When the external Ca^{++} concentration was reduced, the amplitude of the end-plate potentials decreased in a stepwise manner, under various conditions the

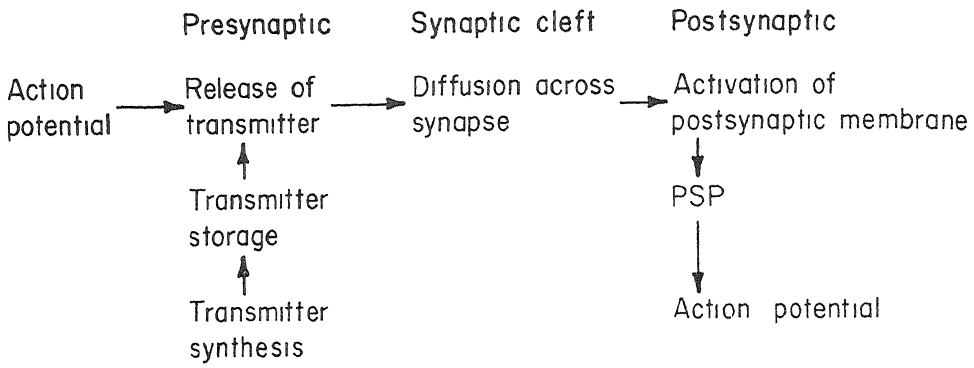


Fig 1 50 Outline of processes involved in storage and release of chemical transmitter

amplitudes were integral multiples of the miniature components However, graded ionophoretic application of acetylcholine directly onto the end-plate produced a continuously graded potential change, thus, indicating that the stepwise electrical changes were not due to the nature of the receptor response, but, rather, to the nature of the transmitter release Their conclusion was that the transmitter is released in quanta or packets of uniform size Depolarization of the presynaptic terminal increased the frequency of quanta release, thus, allowing the MEPP's to summate and reach threshold for firing of an action potential in the postsynaptic cell Other studies have shown that external Ca^{++} is essential for the increased quantal release during the action potential and that external Mg^{++} antagonizes the Ca^{++} action, however, the detailed mechanism for the coupling of excitation to secretion of transmitter is not known

The released acetylcholine diffuses across the synaptic cleft in a very short time (<0.6 msec) to attach to receptor sites on the postsynaptic membrane The mechanism by which this attachment alters the membrane permeability in producing the end-plate potential is not known The action of acetylcholine is terminated by its hydrolysis catalyzed by acetylcholinesterase present at the postsynaptic membrane

Numerous observations indicate that the quantal nature of transmission is applicable to interneuronal synapses in many portions of the vertebrate nervous system and also to other transmitters, for example, norepinephrine In the nervous system, combination of the transmitter with receptor gives rise to the EPSP or IPSP However, direct demonstrations of specific transmitters and their mode of release in the mammalian central nervous system have not yet been achieved

The basis for establishing that chemical transmission occurs involves the following purely chemical criteria A transmitter substance must be found *1*) which is produced presynaptically, that is, the pertinent enzyme systems for its production are present in the presynaptic terminals, *2*) which is released and detected when the presynaptic terminals are stimulated, *3*) the action of which is dissipated by a specific enzyme (as in the case of acetylcholinesterase which destroys acetylcholine) or by reuptake (as in the case of norepinephrine) or by simple diffusion away from the postsynaptic membrane, and *4*) the action of which, when applied directly onto the postsynaptic membrane, mimics the effects of presynaptic nerve stimulation In the case of acetylcholine, the transmitter at the neuromuscular junction, such criteria have been met

Additionally, acetylcholine has been shown to be a transmitter in autonomic ganglia, postganglionic parasympathetic terminals, certain postganglionic sympathetic terminals (sweat glands and vasodilator fibers), and in motoneuron collateral-Renshaw cell synapses in the spinal cord Not all the criteria have as yet been fulfilled to firmly establish acetylcholine as a synaptic transmitter in the vertebrate central nervous system Examples of some of the evidence bearing on the role of acetylcholine as a transmitter in the central nervous system are *1*) choline acetylase and vesicles containing acetylcholine are present in nerve endings from many nuclei of the nervous system; *2*) acetylcholinesterase is abundantly present in neural

tissue, 3) acetylcholine release has been detected from the cerebral cortex after excitation of a sensory nerve, but only in eserinized preparations in which acetylcholine cannot be destroyed, and 4) injection of acetylcholine either directly or electrophoretically will activate many cells in the nervous system

There is substantial evidence that norepinephrine is the transmitter released at postganglionic sympathetic nerve terminals The inactivation of norepinephrine action involves reuptake into the presynaptic terminal Within the central nervous system norepinephrine is found abundantly in the hypothalamus and brainstem reticular formation where it may have a transmitter role On the other hand, dopamine, a precursor of norepinephrine, is abundant in the corpus striatum where it is thought to act as a transmitter Histochemical fluorescent studies have shown that dopamine is present in nerve terminals within the corpus striatum that derive from cell bodies located in the substantia nigra In Parkinson's disease which is associated with symptoms of basal ganglia dysfunction, there is degeneration of this biochemically-defined nigra-striatal path and depletion of dopamine in the corpus striatum Administration of L-dopa, which is decarboxylated to form dopamine in vivo ameliorates the symptoms of Parkinsonism Both catecholamines, dopamine and norepinephrine, are stored in granular vesicles in presynaptic nerve terminals

A number of other substances have been suggested as possible synaptic transmitters and inhibitors Serotonergic axons extend from median raphe nuclei in the brainstem to many areas of the brain, but particularly to the hypothalamus Evidence is accumulating that γ-aminobutyric acid is an inhibitory transmitter in portions of the central nervous system, particularly the cerebellum, and that glycine acts as an inhibitory transmitter in the spinal cord Glutamic acid has been suggested as an important excitatory transmitter It is as effective as acetylcholine in activating neurons when injected into the cells electrophoretically Unfortunately, direct evidence of release has not as yet been obtained for most of the substances alleged to have transmitter or inhibitory activity in the vertebrate central nervous system

There are other indications that chemical transmission takes place at mammalian synapses Some synaptic delay would be anticipated if the release, diffusion, and attachment of a substance were factors important in transmission Thus, the finding of a synaptic delay at vertebrate synapses of 0 5 msec is consistent with a chemical theory A second phenomenon which is consistent with the chemical theory is the fatiguability of synapses when rates of stimulation are increased At 40 or 50 impulses/sec , many synapses within the nervous system fail to transmit an impulse According to the chemical theory, such a failure would be in accord with the expectation that there will be a frequency of stimulation above which the synthesis of transmitter is unable to keep pace with its release from the terminals At such high frequencies, the amount of transmitter released per impulse would decrease and fail to activate the postsynaptic membrane A third, well-known observation is the susceptibility of synapses to anoxia and metabolic inhibitors Thus, cutting off the supply of oxygen interferes with the synthesis of ATP and other chemical materials necessary for the manufacture of the transmitter substance The supply of transmitter diminishes and the amount released per impulse again becomes inadequate to activate the postsynaptic membrane A fourth phenomenon is the repetitive discharge of neurons as noted in the case of the Renshaw cell where a repetitive discharge ensues from a single shock This repetitive discharge is the result of a prolonged depolarization lasting more than 100 msec The prolonged depolarization is attributable to the accumulation of the transmitter substance at the synapse

Criteria for the presence of electrical transmission are essentially the opposite of those for chemical tranmission Thus, a system which transmits impulses across junctions at high frequencies with no delay and in which anoxia and metabolic inhibitors have little or no effect is presumed to be an electrical synapse Such systems behave, indeed, as though no junction were present, although there may be an anatomical junction with little or no space between the presynaptic and the postsynaptic membranes Such electrical junctions have not yet been discovered in the mammalian nervous system but artificial synapses called "ephapses" have been created at junctions between nerve fibers so that an incoming electrical impulse will either modify the excitability in a contig-

uous nerve or actually set off an action potential. Among the invertebrates, several examples of purely electrical junctions have been discovered. The giant motor synapses of the crayfish are designed to transmit electrically in one direction only. Simultaneous recording from both presynaptic and postsynaptic cells has demonstrated that practically no delay (about 0.1 msec.) occurs. The postsynaptic response has the shape of an EPSP and spikes may be generated from sufficiently strong EPSP's. Such synapses appear to act as rectifiers, i.e., they pass current more easily in one direction than the other.

4 / EXCITATION AND CONTRACTION OF MUSCLE

General Characteristics of Muscle Tissue

The functions of muscle tissue are development of tension and shortening Mechanical work is accomplished when the load is not excessive so that the muscle can shorten against it Muscular contraction under physiological conditions depends on the stimulation of muscle tissue through motor nerves In experimental conditions muscular contraction can be elicited by electric or chemical stimulation applied directly to the tissue

Between contractions, muscle tissue assumes a relaxed state characterized by a low resistance to stretch Muscle is elastic but there are changes in its elastic modulus during the transition between the relaxed and the contracted states

The tension developed by a muscle in the body is graded and adjusted to the load If contraction is associated with shortening, the tension is adjusted both to the load and to the velocity of the shortening The graded response of muscle is due to variation in the degree of activation of the tissue through the motor nerves

Through nervous control, the activity of different parts of the muscle tissue of the body is coordinated to produce useful movements and postures The effect of muscle activity is transferred to the skeleton by means of tendons, the translational displacements between various parts of the muscle mass, displacements that are associated with movement, are facilitated by the interposition of connective tissue, either as thin or thick layers of connective tissue or as compact connective tissue septa The latter structures are present at places where the translational displacements are most pronounced

The activity of muscle tissue depends on a supply of energy primarily through breakdown of carbohydrates For highest efficiency and sustained activity, this breakdown is oxidative, which implies a constant supply of oxygen via an efficient blood circulation The transport of oxygen through the blood vessels by means of streaming is supplemented by diffusion from the capillaries through the muscle tissue The dense arrangement of narrow capillaries allowing rapid streaming of the blood secures a fairly short mean distance for the diffusion of oxygen and the maintenance of a high concentration gradient between the capillary lumen and the oxygen-consuming structures, the mitochondria

The activation of the contractile apparatus in muscle tissue depends on events which serve in effect to transfer the nerve stimulus from the motor nerve endings to the contractile machinery Furthermore, the activity of the muscle tissue depends on a precise ionic milieu which is different from that of blood plasma

MICROSCOPIC AND MOLECULAR ORGANIZATION OF MUSCLE

The elongated cells of skeletal muscle (referred to frequently as skeletal muscle fibers or, simply, *myofibers*) are delimited by a plasma membrane, the *sarcolemma* The selective permeability and transport properties of the sarcolemma facilitate the control of the medium in which the contractile machinery is immersed and also contribute to the propagation of impulses that serve to activate the contractile process Skeletal muscle fibers vary in diameter from 10 to 100 μm , and in length from a few micrometers (in the case of insect flight muscle) to several feet (in the case of the long muscles in the giraffe) Despite the great difference in length the mucles of an insect and a giraffe both have the same basic structure Under the light microscope skeletal muscle shows a repeating pattern that has given rise to the designation *striated muscle* Under the electron microscope the subcellular structure of

the skeletal muscle fiber is seen to be composed of smaller fibrous structures 1 μm in diameter *myofibrils*, which are separated by cytoplasm *sarcoplasm*, and arranged in parallel fashion along the long axis of the cell On more detailed examination each myofibril is composed of fibrous substructures, the *myofilaments*, which are of two types, thin myofilaments (approximately 50 Å in width in mammals) composed of a protein called *actin*, and thick filaments (approximately 100 to 120 Å in width) composed of a protein called *myosin*

The arrangement of the myosin and the actin filaments is responsible for the cross-striated appearance of the muscle The cross-striations depend on a regular repetition of one structural pattern consisting of dense cross bands (1 5 μm in width) separated by less dense segments The dense cross bands, referred to as A bands because they are strongly anisotropic, contain the thick myosin filaments arranged neatly in parallel (Fig 1 51) When myosin is extracted from intact muscle, thick filaments can no longer be detected microscopically

The myosin filaments have a characteristic appearance which reflects the mode of aggregation of the myosin molecules within the filaments With the exception of a middle section of the filaments, a series of lateral projections extends from the approximately 100-Å-thick trunk of the filaments Similar filaments can be made artificially by allowing myosin in solution to aggregate Under such conditions the gradual growth of filaments is reflected in the distribution of sizes of reconstituted filaments The morphology of the filaments agrees with the assumption that the myosin molecules aggregate with their globular ends directed towards the ends of the filaments The middle smooth shaft corresponds to a close packing of the rod-shaped part of the molecules and the series of projections corresponds to their globular ends (Fig 1 52)

Such a polar arrangement of the myosin molecules could secure a definite direction of the force developed in connection with a configurational change of the myosin molecules whether the latter change is minimal or extensive

The smooth shaft of the filament shows a somewhat thicker segment located exactly in the middle of the filament, contributing to a dense line, the M band, extending across the middle of the sarcomere

The actin filaments in intact muscle show

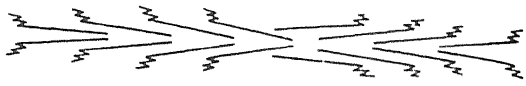


Fig 1 51 Schematic presentation of the likely arrangement of myosin molecules in the myosin filaments

great similarities to fibrous actin (F actin) filaments that are formed in vitro by polymerization of globular actin (G actin) although the axial periodicity is slightly greater (406 to 411 Å) for the natural than for the artificial filaments (350 Å) Recent analyses of electron micrographs of F actin and isolated thin filaments indicate the presence in the latter of additional proteins This additional material probably reflects the presence in the thin filament of the tropomyosin troponin complex as does the 406 to 411 Å periodicity (vide infra)

The less dense segments the *I bands*, contain the actin filaments which extend symmetrically in opposite directions from a dense thin line, the *Z line* (Fig 1 53A) The Z line appears in embryonic muscle before the actin filaments develop and is comprised of a lattice like protein This protein is intertwined with the actin (thin) filaments which extend perpendicularly from the Z line for about 1 μ on either side When actin is extracted from intact muscle the thin filaments can no longer be observed microscopically The term I band is based on the fact that this zone was originally considered to be isotropic Although it is now recognized to be weakly birefringent, it is very much less birefringent than the A band. The width of the I band varies with the degree of stretch or shortening of the muscle fiber Since the length of the individual actin filaments is about 1 μm and there are two sets of such filaments extending from the Z line, the total width of the Z line-actin filament complex will be about 2 μm The Z-line structure contributes toward keeping the actin filaments arranged in register and with a regular spacing The gap between the terminations of the actin filaments is called the H zone and, as noted above, the M line is the darker area in the center of the H zone The M line is particularly pronounced during contraction

The term *sarcomere* defines the repeating unit of this striated structure as the region between two consecutive Z lines, thus, this unit consists of one A band and one-half I band at each end of the A band The sarcomere is the



Fig 1 52 Electron micrographs of negatively stained preparations of (*left*) natural thick filaments prepared directly from homogenized muscle (*middle*) synthetic filament formed by aggregation of purified myosin in 0 1 M KCl and (*right*) synthetic filament formed by aggregation of light meromyosin (vide infra) in 0 1 M KCl Projecting cross-bridges can be seen on the natural and synthetic filaments in which whole myosin is present—that is in which both the heavy- and light-meromyosin parts of the myosin molecule are represented The filaments containing light-meromyosin alone have no projections (about × 162 000) (Huxley 1969)

fundamental contractile unit of muscle Its length at rest varies from 10 to 15 μm in crab muscle and between 2 and 2 5 μm in frog and mammalian (including human) muscle

At the rest-length of muscle, which here is defined as the normal length of the muscle in situ, the two sets of filaments in most muscles interdigitate with a rather extensive zone of overlap at each end of the A band In the rabbit psoas muscle the sarcomere length at rest-length is about 2 8 μm If the muscle is stretched, the zone of overlap decreases in width in proportion to the increase in half-width of the I band The zone in the middle of the A band bounded by the two zones of overlap, the H zone, also increases in width, this increase corresponding to the decrease in width of both zones of overlap This means that when a muscle is stretched, the actin filaments are being pulled out of the A band, while the length of both the myosin filaments and the actin filaments remains constant (Fig 1 56 *B* and *C*) At a sarcomere length of about 3 5 μ, the actin filaments have been pulled out completely, and the two sets of filaments appear now in an end-to-end arrangement At further stretching, the two sets of filaments become separated by a gap, in which filaments consid-

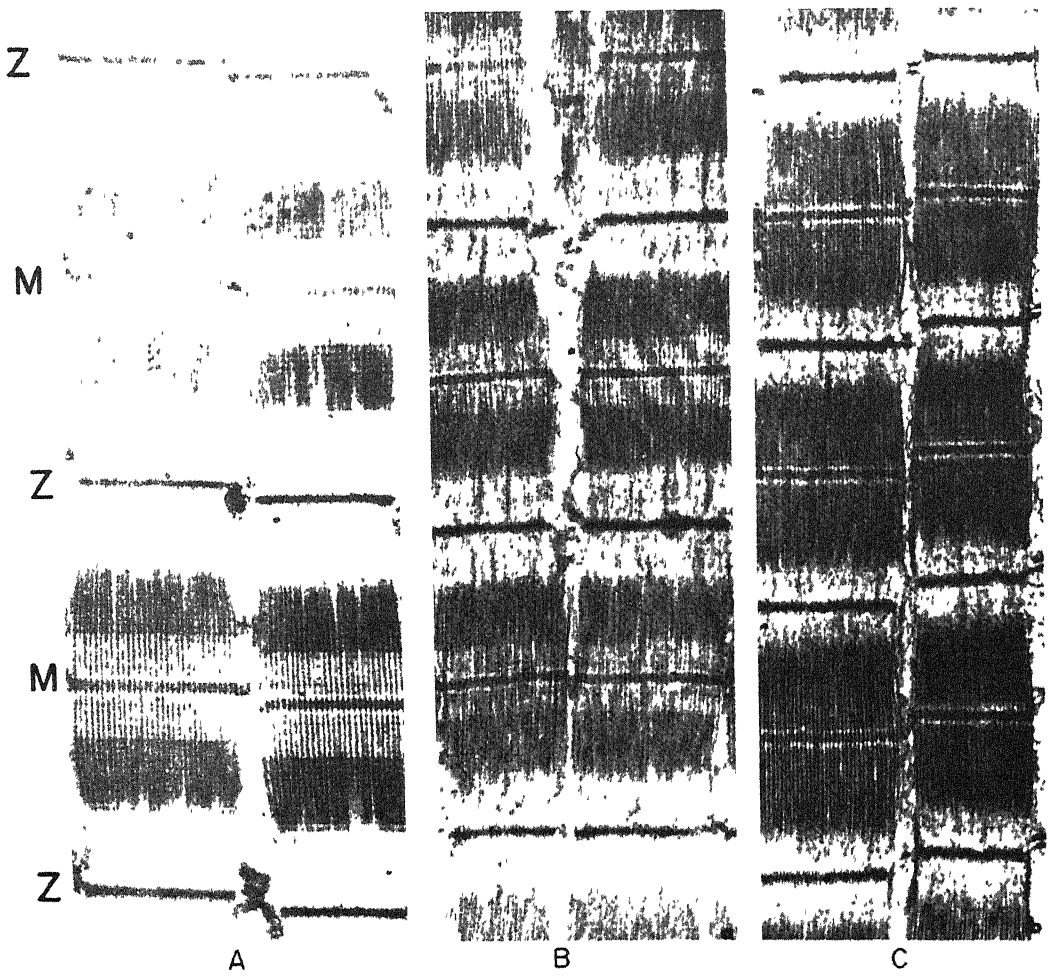


Fig 1 53 *A* Frog sartorius muscle (thick section) showing periodicity in both thin filaments (as seen in I bands) and in thick filaments (as seen within the entire width of the H-zone) Periodicity is best seen by viewing pictures of the fibrils obliquely from the side The two periods differ by 5 to 10% (their probable values being 410 Å and 435 Å) Near the edges of the H-zone the periods show strongly in the overlap region where they are more or less in register but as they get progressively further and further out of register the vernier effect blurs out the pattern The M line structure (3 dense lines, about 200 Å apart running across the fibril and probably corresponding to the cross bridges between thick filaments seen in the M-region in cross-section) shows up rather clearly The less dense region immediately on either side represents the remainder of the central region of the thick filaments from which the cross bridges are absent (× 32 400) *B* Frog sartorius muscle (live) fixed in gluteraldehyde during direct electrical stimulation and whilst it was shortening against a load postfixed in osmium tetroxide stained with PTA in absolute alcohol, araldite embedding Only surface fibers used Specimen shown had only shortened very slightly, and quite a wide H-zone is still visible (× 19 200) *C* As in *B* but allowed to shorten further so that a very short H zone (and a shorter I band) is visible (× 19 200) (Huxley 1964)

erably thinner than the actin filaments can be seen extending between the ends of the myosin and the actin filaments

If a muscle is allowed to shorten, either passively or through an active contraction, and the structure is analyzed after the muscle has acquired a new length, the relationship between the half-width of the I bands, the width of the overlap zone and the width of the H band is the same as described above At sarcomere lengths less than 2 μ, that is, less than the sum of the lengths of the two sets of actin filaments in the sarcomere, a somewhat more dense zone appears in the middle of the A band This phenomenon is due to a double overlap of the actin filaments (Fig 1 56 D)

When the sarcomere length has decreased to 1 5 μ or less, the Z lines cause the ends of the thick filaments to crumple, giving rise to zones with irregularly arranged filaments Such zones are referred to as contraction bands Contraction bands at this sarcomere length involve a rather random arrangement of thin actin filaments as well as thin filaments derived from the myosin filaments as a consequence of the splitting of the latter into filamentous subunits

In accord with the above considerations the microscopic patterns seen in cross sections of muscle fibers depend on the level of the section (Fig 1 54) For instance, near the Z band, only actin fibers are observed (Fig 1 55*A*) whereas at the site of overlap of actin and myosin filaments, each thick filament is surrounded by six thin filaments and each thin filament by 3 thick filaments (Fig 1 55*B*)

The sarcolemmal membrane of the myofibers has the general double layer structure of most biological membranes Besides shielding the muscle fiber, the sarcolemma imparts a characteristic resistance to stretch for the muscle as a whole Because this resistance is anatomically disposed in parallel with the contractile elements, it is usually lumped together with the resistance contributed by the connective tissue surrounding the muscle fibers and both are referred to as *parallel elastic elements*

Skeletal muscle fibers are multinucleated In embryonic tissue the nuclei are located in the center of the fiber, however, in the course of differentiation they come to occupy a peripheral position

Skeletal muscle has a rich supply of mitochondria and microsomes dispersed along the length of the myofiber These subcellular structures, as in other tissues, are supportive components of the muscle fiber, the myofiber is the smallest metabolically autonomous unit

PROTEINS OF THE CONTRACTILE ELEMENTS

1) **Actin.** With recent advances in the purification of muscle constituents it has become apparent that actin comprises a major portion but not all of the thin filament (I band) Actin molecules exist in two states G-actin, a globular monomer, which is stable in solutions of low ionic strength, and F-actin, a fibrous polymer, which is formed when salts are added to G-actin The polymerization of G-actin to form F-actin (referred to as the G-F polymerization) is accompanied by an increase in the viscosity of the protein solution The conformation of the F-actin polymer is that of a double-stranded helix

A bound nucleotide and a bound divalent cation are necessary for polymerization of globular actin In the in vitro preparations, 1 mole of adenosine triphosphate (ATP) and 1 mole of

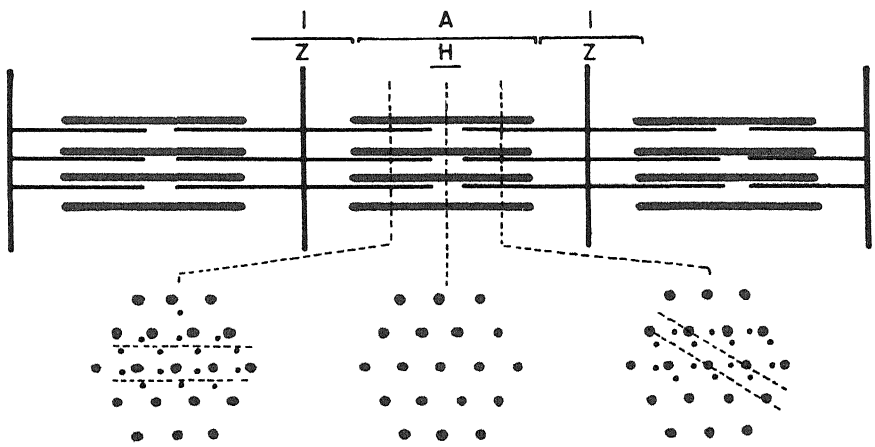


Fig 1 54 Diagram illustrating the arrangement of the two kinds of protein filaments in a myofibril At the top are three sarcomeres drawn as they would appear in longitudinal section Below are transverse sections taken through the H zone (center) and through the other parts of the A band (right and left) where the arrays of thick and thin filaments interdigitate If a longitudinal section were cut in the plane which is indicated by dotted lines in the left-hand transverse section it would appear as in the electron micrograph in Figure 1 52 and as in the diagram here, and it would show (apparently) one thin filament between each two thick ones If it were cut in the other plane (dotted lines in right-hand transverse section) it would show two thin filaments between each two thick ones (From Huxley and Hanson 1960)

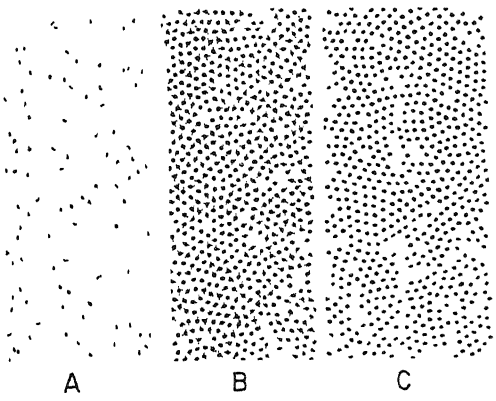


Fig 1 55 *A* Cross section of muscle fiber near Z band showing only actin fibers *B* In region of overlap both thick and thin filaments are seen *C* In the H zone only myosin filaments appear

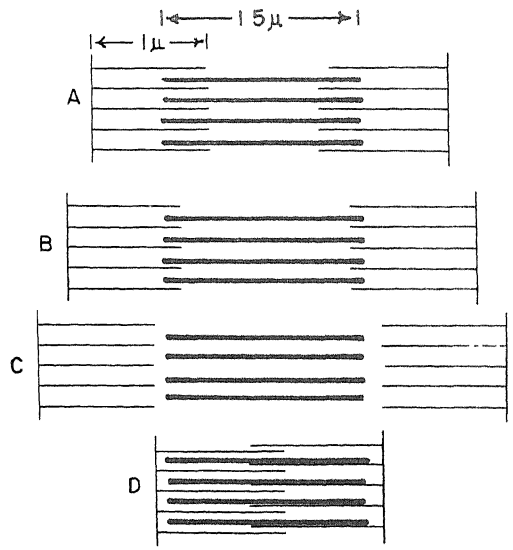


Fig 1 56 Schematic illustration of the arrangement of thick myosin filaments and thin actin filaments in the sarcomere at various degrees of stretch and shortening *A* and *B* show different degrees of overlap of the two types of filaments *C* illustrates the case when the actin filaments have been completely pulled out from the A band with a gap between A band and the actin filaments In the extensively shortened sarcomere in *D* a zone of double overlap of actin filaments has developed in the center of the A band

Ca^{++} have been found to be associated with 1 mole of G-actin Nucleotides other than ATP can also bind with G-actin, although with lower affinity Mg^{++} and Mn^{++} can replace Ca^{++} but also with the affinity that is four times less than the affinity of Ca^{++} for actin

The nucleotide (ATP) bound to G-actin is readily exchangeable, however during polymerization to F-actin, the terminal phosphate of ATP is split off and the resulting adenosine diphosphate (ADP) formed remains bound to F-actin and is not exchangeable Addition of inorganic phosphate (P_i) per se will not reverse the reaction of G-F polymerization to restore G-actin and ATP in fact such a depolymerization requires ATP

The molecular weight of skeletal G actin is generally accepted to be approximately 47 000 based on viscosity measurements, sedimentation equilibrium studies, amino acid analysis, and ATP content

2) **Tropomyosin.** Tropomyosin is a protein which constitutes 10 to 12% of skeletal myofibrillar protein Although it was purified in 1948 its role in the machinery of the contractile process had escaped attention until recently It is now believed to react with troponin to form a complex which is implicated in the processes by which excitation is coupled to contraction The molecular weight of tropomyosin is approximately 70,000 Tropomyosin is a coiled coil (two alpha helices wound around each other), which by itself has no known biological activity When bound to actin, however, tropomyosin can modify the interaction between actin and myosin

The presently available evidence does not support a previous suggestion to the effect that tropomyosin comprises the structural protein of the Z line Antibodies against tropomyosin show very weak binding to the Z-line proteins and ammonium sulfate extracts of tropomyosin fail to restore the Z band to rabbit muscles from which the Z line had been previously extracted with glycerin

3) **Troponin.** This protein was isolated recently It has been studied extensively because it is linked with the important regulatory action of calcium in the contractile process Its salient feature is a high affinity Ca^{++} binding site, 2 moles of Ca^{++} are bound to approximately 100,000 g of troponin Recent work indicates that troponin consists of two or three distinct components, one of these, designated as troponin A, possesses the specific (high affinity) binding site for Ca^{++} Another component, designated as troponin B, inhibits the Mg^{++}-activated interaction of actin and myosin which constitutes the quintessential biochemical event in the contractile process (vide infra) Troponin molecules complexed with

tropomyosin are believed to be distributed along the thin (actin) filaments at intervals of 406 to 411 Å to form reactive sites which are important in the contraction process Antibodies to troponin are not found on the Z band and it may therefore be concluded that troponin, like tropomyosin, is probably limited only to the thin filaments

4) **Myosin** This protein is the chief constituent of the thick filaments which, as has been noted, disappear after myosin has been extracted from muscle fibers When purified myosin molecules are allowed to aggregate, the resulting polymers are virtually indistinguishable from the thick filaments Myosin is a large protein with a molecular weight between 450,000 and 500,000

The physicochemical properties of myosin indicate a high degree of molecular assymmetry One end of the molecule is globular, the other end is rodlike Optical rotatory dispersion measurements indicate that between 54 to 60% of the peptide chain in the native molecule exists in the α helical configuration The extremely large size of the myosin molecule suggested a complex structure, and indeed it was found that treatment of myosin with proteolytic enzymes such as trypsin splits the protein into two components, light meromyosin (LMM), which consists of most of the rodlike part of the native molecule, and heavy meromyosin (HMM) the primarily globular end of the native molecule, to which is attached a short piece of the rodlike portion

Myosin possesses adenosine triphosphatase (ATPase) activity which is associated with the HMM component, LMM being totally inert in this respect

Unlike trypsin, papain cleaves myosin at a point of the globular end of the molecule splitting off two identical HMM components These HMM components are held together and attached to LMM by a peptide stem at a locus which is not susceptible to the action of trypsin (Fig 1 57) When myosin is treated with 12 M urea or 5 M guanidine, it can be resolved into two heavy fragments and two or three light fragments The functions of the latter are unknown, however, differences in electrophoretic mobilities of light myosin fragments, when they have been prepared from different muscle types, suggest but do not prove that the nature of the light myosin fragments may determine myosin ATPase activity

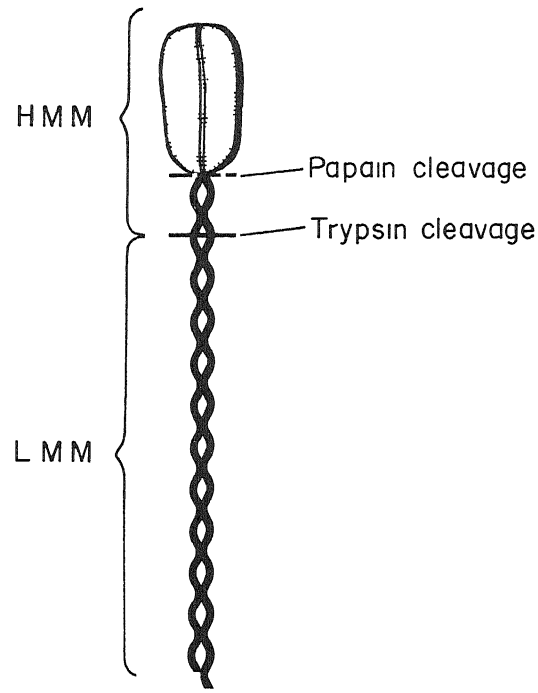


Fig 1 57 Sites of cleavage of the myosin molecule by papain and trypain

LMM exists in a highly ordered structure Like tropomyosin, LMM consists of two coiled parallel peptide chains each in the form of an α helix HMM, on the other hand, exists primarily in globular form with a short highly helical fibrous rodlike extension Acid treatment of HMM releases both the rod and the globular head of the molecule, the latter in the form of its two subunits each of which retains ATPase activity and the ability to combine with actin

Myosin can be split into subunits by 12 M urea and 5 M guanidine (as noted above) and also by other agents such as alkali and succinic anhydride The light subunit appears to be associated with the ATPase activity and the species specificty of myosin However, variants of the whole myosin molecule with characteristic amino acid sequences can be found in different muscles of the same animal and also in the same muscle at various developmental stages (vide infra)

The molecular weight of skeletal HMM is approximately 350,000 and that of LMM is approximately 150,000 The estimated molecular weight of the heavy fragments of myosin is between 112,000 and 180,000 whereas the

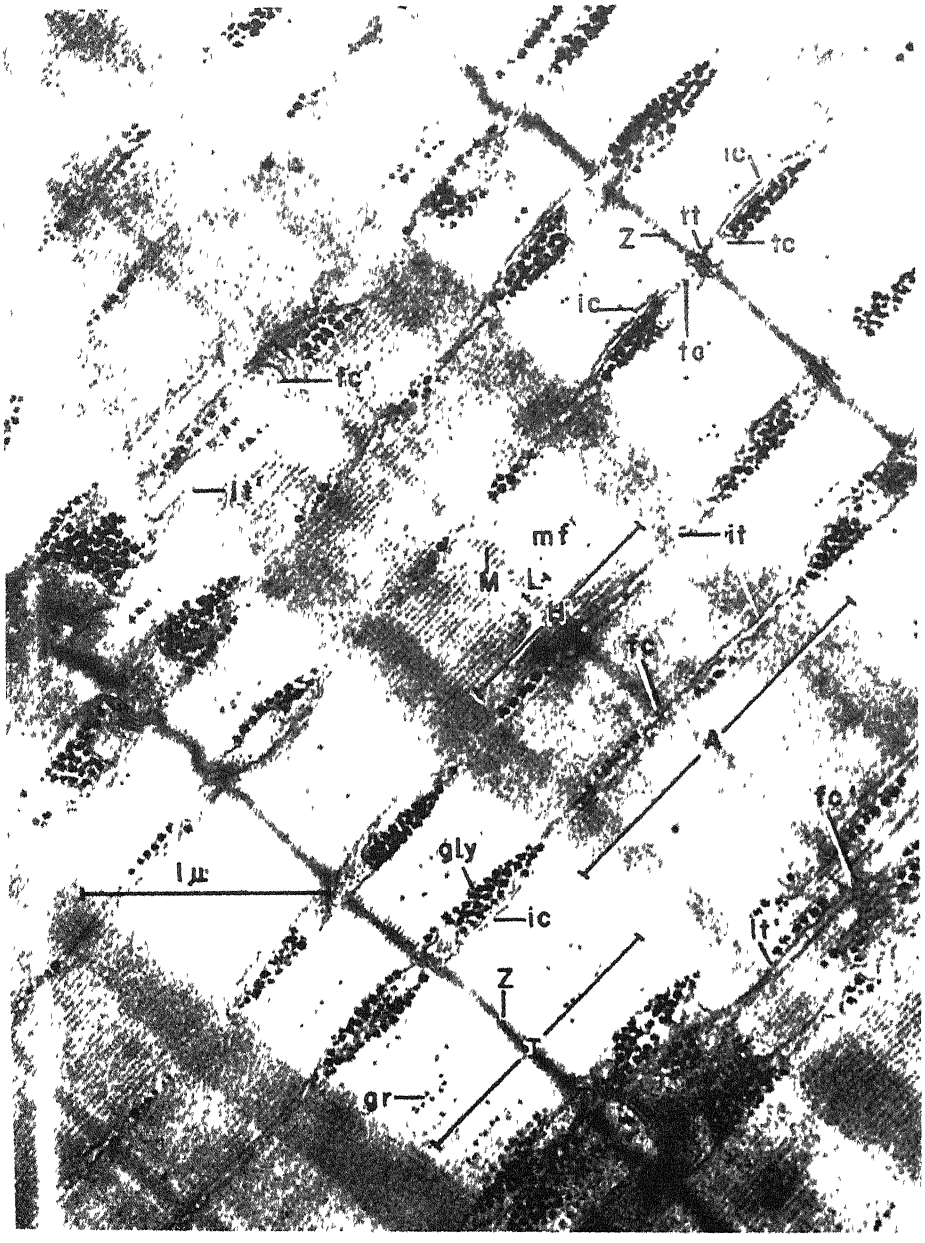
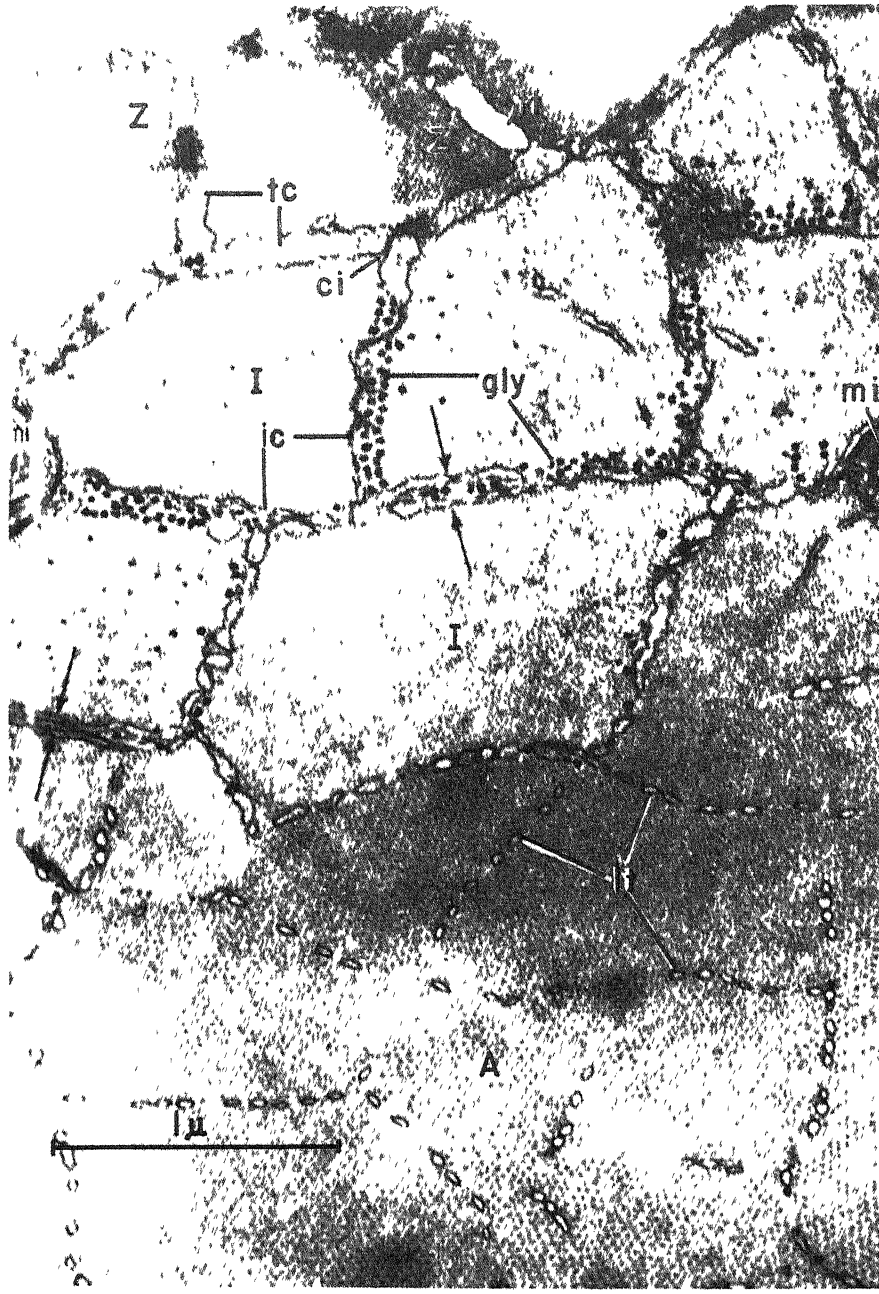


Fig 1 58 ***A*** Longitudinal section oriented with the long axis of the fiber extending from lower left to upper right in the figure The myofibrils (*mf*) show a typical transverse banding pattern with dense Z lines (*Z*) spaced about 3 μ apart at the centers of the light I bands (*I*) The A bands (*A*) have a central light H zone (*H*) and denser outer regions where primary and secondary filaments overlap At the center of each A band is a light L zone (*L*) with a dark M line (*M*) down its middle Triads consisting of two terminal cisternae (*te*) and a centrally located transverse tubule (*tt*) are seen adjacent to the Z lines (*Z*) of the myofibrils Connecting in the terminal cisternae are the intermediate cisternae (*ic*) Adjacent to the myofibrillar A bands elongated profiles of the longitudinal tubules (*lt*) and oval profiles of the fenestrated collar (*fc*) are seen These latter structures are also seen overlying the myofibrils at two places where the section grazes the surface of the myofibril (*fc'* and *lt'*) The large, densely stained granules are glycogen (*gly*) Smaller granules (*gr*) are found within the I bands (× 36 000) ***B*** Very thin, slightly oblique transverse section showing a Z line (*Z*) as dark patches with a square pattern at the upper left corner of the figure The I band (*I*) appears as the lighter areas around and below the Z line patches The A band (*A*) is at the bottom of the figure Near the Z



line and for about one-third of the I band the fibrils are completely surrounded by wide membrane-bounded elements of the triads Dark patches within these elements are interpreted as regions where the membranes of the central portion of the triad (transverse tubule and flattened ends of terminal cisternae) lie within the section The triad elements that are filled with granular material are the terminal cisternae (*tc*) and the very light patches (*tt*) are interpreted as areas where only the transverse tubule lies within the section, implying that the section thickness in this region is somewhat less than 300 A One interruption in lateral continuity of the terminal cisternae is indicated (*ci*) Below these elements in the figure but still within the I bands are found the narrower profiles of the intermediate cisternae (*ic*) These are found alongside and sometimes flanking both sides of the glycogen (*gly*) masses (see arrows) The intermediate cisternae extend as far as the edge of the I band Beyond this within the A band roughly circular profiles of the longitudinal tubules (*lt*) are seen Parts of two mitochondria (*ml*) appear in this figure (× 45 000)

myosin light fragments have molecular weights of approximately 20 000

5) Other Proteins of the Contractile Machinery In addition to actin and myosin and their derivative subunits, other proteins have been isolated from myofibrils One of these proteins appears to have been derived from the M line Another, of unknown derivation, modifies the aggregation of myosin However, at the present time the role of these proteins in the contraction-relaxation cycle of muscle is not known

A protein designated as α-actinin often isolated with actin, is found to increase the ATPase activity of myosin When bound to F-actin, the α-actinin-actin complex forms a gel The significance of α-actinin in muscular contraction however, remains in dispute

α-actinin has three distinct components, the lightest of which (6S component) has a molecular weight of 160,000, has been localized on the Z lines and can be displaced by troponin

β actinin, another protein associated with the contractile system of muscle, is probably an important determinant of the structure of the thin filament inasmuch as it has been shown to *1*) bind stoichiometrically to F-actin, *2*) inhibit the network formation of F-actin, and *3*) induce the formation in vitro of 1 μ filaments similar to actin filaments

THE SARCOPLASMIC RETICULUM

In 1902, Veratti observed a lattice or network structure in the muscle cytoplasm (sarcoplasm) similar to what is now recognized as the endoplasmic reticulum Veratti proposed that this newwork may be a continuation of nerve endings which function to convey the excitation process to the contractile elements These prescient observations and suggestions were ignored for over 50 years until the recent era of detailed, critical study of cellular ultrastructure

The ultrastructure of the sarcoplasmic reticulum (Fig 1 58) was first described by Bennett (1955) and by Porter and Palade (1957). They observed an elaborately anastomosing tubular network running parallel to the myofilaments The diameter of this tubular system was approximately 5 μ The slender tubules often extended the full length of the sarcomere and were closed at each end (terminal cisternae), the separation of two such adjoining cisternae taking place at the level of the Z line and amounting to 500 Å A third vesicular structure the transverse or T tubule was also identified in the gap between two adjourning terminal cisternae This *triad* of structures T tubule and two terminal cisternae were observed to repeat regularly in the I band at the level of the Z line or at the junction of the A-I band or at both sites depending on the type of muscle studied and the species from which it was obtained The regularly repeating pattern of the network (Fig 1 59) suggested from the outset that the sarcoplasmic reticular system in general and the triad in particular were intimately related to the function of the contractile elements

Recent studies of the embryonic development of muscle have shown that the sarcoplasmic reticulum (SR) originates from the nuclear membrane before the differentiation of myofila-

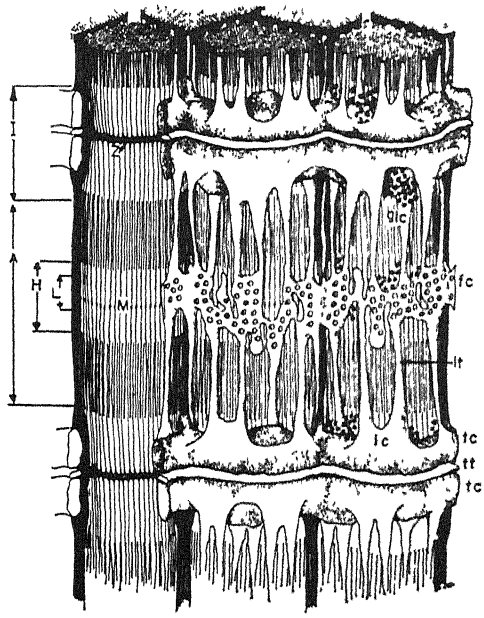


Fig 1 59 A three dimensional schematic diagram illustrating the sarcoplasmic reticulum and its relationship to the myofibrils The myofibrils show the light band (*I*) with the dense Z line (*Z*) at its center The A bands (*A*) have a central light H zone (*H*) and denser outer regions where primary and secondary filaments overlap At the center of each A band is a light zone (*L*) with a dark line (*M*) in its center Triads consisting of two terminal cisternae (*tc*) and a centrally located transverse tubule (*tt*) are seen adjacent to the Z lines of the myofibrils Connecting to the terminal cysternae are the intermediate cysternae (*ic*) and adjacent to the A bands they form the longitudinal tubules (*lt*) In the vicinity of the H zone in the center of A band the longitudinal tubules join the fenestrated collar (*fc*) *glc* glycogen granules (From Peachey 1965)

ments takes place and also that the SR lays the foundation for formation of the Z lines Other lines of evidence indicate that the central component of the triad (the T tubule) is an invagination of the sarcolemmal membrane which penetrates deep inside the fiber (Fig 1 60) For instance, it has been shown that if muscle is incubated with ferritin, the granules can be later found inside the T tubules It was also shown by Huxley and Taylor in frog and fish skeletal muscle, that following exploratory stimulation with minute electrical currents applied with a microelectrode at intervals of 1 μ along the length of the sarcolemmal membrane, a local contraction was elicited *only* when the stimulatory microelectrode was placed at the level of the Z line

In a similar experiment in crab muscle, a local contraction was elicited only at the A-I junction which, *in this species*, is the site in the sarcomere at which the T tubules are located These findings revealed an important relationship between the T tubules and the adjoining cisternae of the triad and also between the triad as a whole and the contractile elements (myofilaments) of the myofiber

More recently it has been shown *1*) that the T tubules of frog skeletal muscle can be shattered and converted into vesicles by soaking the tissue for about 2 hr in hyperosmotic Ringer's solution containing 20 to 50% glycerol and then subjecting the tissue to osmotic shock by returning the tissue abruptly to normal Ringer's solution, and *2*) that following this disruption of the T system it was not possible to elicit a contractile response even though an action potential could be recorded after electrical stimulation Thus, it was demonstrated that an in-

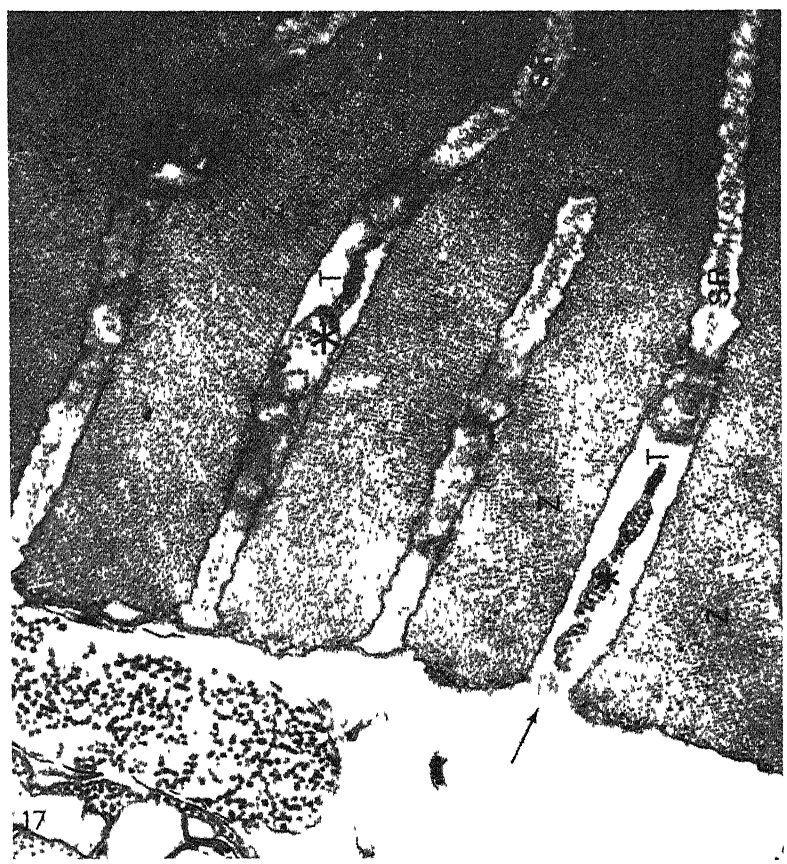


Fig 1 60 Transverse section of a glutaraldehyde-fixed muscle fiber to show the form and disposition of the T system The section cuts across the Z line (*Z*) and so includes the middle of the T system cisterna Openings or orifices of the T system (*T*) to the outside are shown at the arrow Strands of sarcoplasm (asterisk) which penetrate into the T system from neighboring fibers may be observed *SR* sarcoplasmic reticulum Original magnification × 38 000 (From Franzini-Armstrong and Porter 1964)

tact T tubular-triadic system is necessary for the electrical excitation of muscle to become translated into a contractile response

Theories of Contraction

FOLDING PROTEIN THEORY

There have been two concepts regarding the structural changes occurring during shortening According to an older concept, the contraction of muscle is associated with a splitting up of the compact myosin filaments at their ends into a number of subunits possibly consisting of individual myosin molecules, an interaction of these subunits with the actin filaments resulting in bonds being formed between the subunits and the actin, yielding drastic changes in the myosin subunits from the originally extended configuration to a compact configuration This would involve extensive conformational changes in the myosin molecules and would result in a pulling in of the actin filaments toward the center of the A band since the myosin subunits remain attached to the myosin filaments at the end which is located toward the center of the sarcomere The splitting up of the end regions of the myosin filaments would be an important first step since it would allow active sites on the myosin molecules to be exposed In the compact arrangement of myosin in the thick filaments these sites would be shielded from interaction with actin

THE SLIDING FILAMENT THEORY

According to the other concept of muscle contraction, the "sliding model," there are no extensive configurational changes occurring during contraction The actin filaments slide along the surfaces of the myosin filaments due to oscillatory, minute movements of the lateral process of the latter filaments with repeated alternative formation and breaking of bonds between myosin and actin The bonds are formed in the oscillatory cycle when the processes are in a position closest to the Z line, the bonds are broken when the processes have swung to their extreme position furthest from the Z line and closest to the center of the A band

The first evidence that led to the sliding filament theory came from electron microscopic observations which showed that cross-bridges link the actin and myosin filaments together across a gap of about 130 Å The bridges are seen as projections on the myosin filaments even in places where no actin filaments lie alongside them (for example in the H zone of stretched sarcomeres) but they do not project from the actin filaments in the portion of the I band which does not overlap with the thick (myosin) filaments Therefore it seems clear that the cross-bridges are permanent parts of the myosin filament structure On the basis of the above observations it was suggested that cross-bridge structures were the mechanical links between the thick and thin filaments It was also suggested that cross bridges probably represented the heavy meromyosin subunits of the myosin molecule Because the ATPase activity and the actin binding properties of myosin reside on the heavy meromyosin it is reasonable to regard this enzymatic activity as a built-in feature of the structure of the cross-bridges and to assume that the cross bridges have a special affinity for sites on the actin molecules

During a twitch, the frog sartorius sarcomere, which typically has a length of 2.5 μ, shortens by about 30%, the actin and myosin filaments sliding past each other by 0.375 μ in each half sarcomere

Huxley (1964) observed that variations in orientation of the cross-bridges can be seen in electronmicrographs of resting and contracted muscle The distal ends of the cross bridges never seem to be displaced by more than about 100 Å from the position they would occupy if the bridges were exactly perpendicular to the thick filaments Therefore, he proposed that, in order to produce the much larger overall sliding movement, some type of repetitive interaction of the cross-bridges with the actin filaments is necessary and he suggested two possible mechanisms for this interaction *1)* the cross-bridges move to and fro in a cyclic manner, first attaching to the actin filaments and pulling them toward the center of the A band in the initial phase of the cycle and then detaching again prior to their return stroke—similar to the action of a ratchet, *2)* the cross-bridges might remain rigidly fixed in position on the myosin filament while repetitive cyclic changes occur at fixed intervals along the actin filaments causing the latter to attach and then detach from the fixed cross-bridges, thus, moving to the center of the sarcomere (H zone) In either case, the central idea is that

the cross-bridges are attached to actin filaments while force is being developed and the cross-bridges serve as the agents through which the mechanical force is transmitted

From the known free energy of hydrolysis of ATP, it seems possible that splitting of 1 molecule of ATP per cross-bridge can provide enough energy for the actin and myosin filaments to slide past each other for a distance of 50 to 100 Å during a maximal shortening This is consistent with a model in which 1 molecule of ATP is split at each cross-bridge during one tension-generating cycle and in which each cross-bridge can go through its cycle only once during a relative movement of the filaments over a distance of 50 to 100 Å

In this context, there will be neither tension development nor significant energy release (ATP-cleavage) when this "make and break" filament cycle is not completed Thus, chemical energy release in muscle can be controlled *1)* by the tension developed (which is proportional to the number of cross-bridges that had time to attach themselves to the actin filaments at any given shortening velocity), and *2)* by the distance shortened (which would be proportional to the number of "make and break" cycles) This kind of system would release energy efficiently and in proportion to the work done

Strong evidence supporting the notion that the cross-bridges represent the heavy meromyosin end of the myosin molecules came from a comparison of the filamentous aggregates of light meromyosin (formed at physiological ionic strength) with aggregates formed under similar conditions by intact myosin molecules (Fig 1 52) The light meromyosin filaments were seen to consist of needle-shaped structures many microns long and of variable widths up to several thousand angstrom units, the surfaces of these filaments were perfectly smooth On the other hand, the filaments formed by aggregation of whole myosin molecules had a large number of projections on their surfaces over most of their length These latter filaments varied in thickness but were usually less than 200 Å in diameter and shorter than the light meromyosin filaments These synthetic myosin filaments were very similar in their appearance to the natural myosin filaments prepared from disrupted muscle It is therefore apparent that the cross-bridges observed by negative staining methods are equivalent to the cross-bridges in the sectioned material and, since they do not appear in the light meromyosin preparations noted above, they must be associated with the heavy meromyosin part of the molecule Their dimensions closely approximate those of the sectioned cross-bridges (40 to 50 Å by 120 Å) if shrinkage (ca 20%) of the sectioned preparations is taken into consideration

The width of isolated heavy meromyosin (determined by shadow casting technique) is about 40 to 50 Å in the globular region which is attached to a short tail approximately 400 Å long From the above observations it is concluded that the cross-bridges consist of the globular head of the heavy meromyosin, supported by an α-helical tail which lies in an approximately parallel position along the backbone of the myosin filament to which it ultimately attaches Myosin molecules are arranged in the thick filaments with a definite structural polarity so that the heads of the molecules are always directed away from the midpoint of the filaments Thus, all cross-bridges in one half of an A band have the same orientation (polarity) This polarity is reversed in the opposite half of the A band The actin monomers are also oriented oppositely on either side of their attachments to the Z line Thus, during contraction the sets of actin filaments in each half sarcomere are drawn toward the center of the A band and subjected to sliding forces oriented in opposite directions This polar organization of both the myosin and actin filaments is a fundamental condition of sliding filament theory

Accepting the evidence that the cross-bridges attach to active sites on the actin filaments in order to develop force, it is reasonable to link this interaction to the hydrolysis of ATP and the release of energy The fact that actomyosin has considerably greater ATPase activity than does myosin alone supports this idea It is not known whether ATP is split before, during, or after the combination of actin with heavy meromyosin (cross-bridges), however, there is no doubt that this force-generating interaction requires the transformation of energy from ATP into mechanical work Conversely, it is important to recognize that the chemomechanical transformations of muscle also *control* the release of energy from ATP (Fenn effect, vide infra)

The contribution of each cross-bridge to the force generated by contracting muscle will be considered in the discussion of the length-tension relationship At first it was assumed that cross-bridges undergo specific structural

changes when they interact with actin and split ATP However, measurements of the spacing between actin and myosin at different sarcomere lengths revealed that the filaments move closer together as the muscle is stretched and further apart when the muscle returns to rest length, the separation varying inversely with the square root of the muscle length Thus, as the sarcomere shortens from 2 8 to 2 0 μ, the side spacing increases by about 18% of its original value In absolute terms it has been estimated that the center to center distance between actin and myosin filaments may increase by 40 Å as the sarcomere shortens from 2 8 μ to 2 0 μ For contact to be made between actin and myosin filaments during contraction, the cross-bridges must adapt to these changes in spacing between the filaments without any qualitative alteration in function

The technique of negative staining has helped to elucidate some of the puzzling aspects of the actomyosin interaction For example, this technique has revealed significant irregularities in the repeating subunit of the actomyosin filaments The double helical structure of G-actin units remain clearly visible (Fig 1 61), whereas, no trace of an ordered structure (helix) can be seen in the projection of the myosin filaments However, when heavy meromyosin is reacted with actin, a clear regularly repeating pattern of arrowhead-like structures bound to the actin filament can be observed (Fig 1 62), these arrowhead-like structures represent myosin bound to the sites on actin at which the cross-bridges link It is not known why the sites of attachment of the head of the cross-bridges on the thin (actin) filaments are more orderly than the sites of attachment of the tail of the cross-bridges on the thick (myosin) filaments, however it is possible that the cross-bridges are more rigidly bonded to the actin filament than to the myosin backbone

Using x-ray diffraction Huxley had initially predicted that the contractile components (later identified as actin and myosin) must have different structures Further study of x-ray diffraction patterns indicated that in striated muscle of vertebrates the cross-bridge projections on the thick filaments are arranged on a two-stranded helix with three pairs of cross-bridges per 360° turn of the helix (6/2 helix) Each pair of bridges project out from the myosin backbone in opposite directions so that the two bridges in any pair form an angle of 180° with each other, i e , any pair of cross-bridges form a straight line at any given level. Taking any pair of cross-bridges as a reference, the nearest neighbor pair of bridges on either side occurs at a distance of 143 Å and is rotated relative to the reference pair by 120° This arrangement continues so that the full pattern repeats itself at intervals of 3 × 143 Å or 429 Å. (Fig 1 63)

On the basis of these observations a model

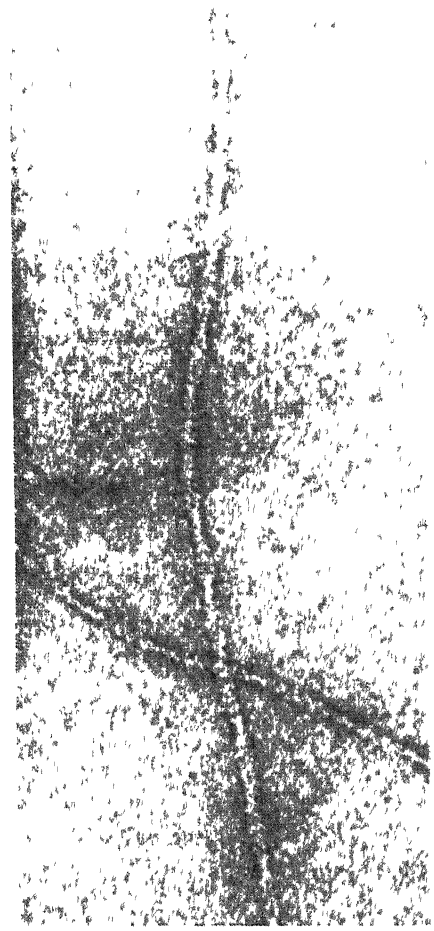


Fig 1 61 Electron micrograph of negatively stained actin filament, showing the double helical arrangement of two chains of globular subunits twisted around each other The subunit repeat in each chain is about 55 Å and the cross-over points of the two chains are 360 to 370 Å apart (Huxley, 1969)

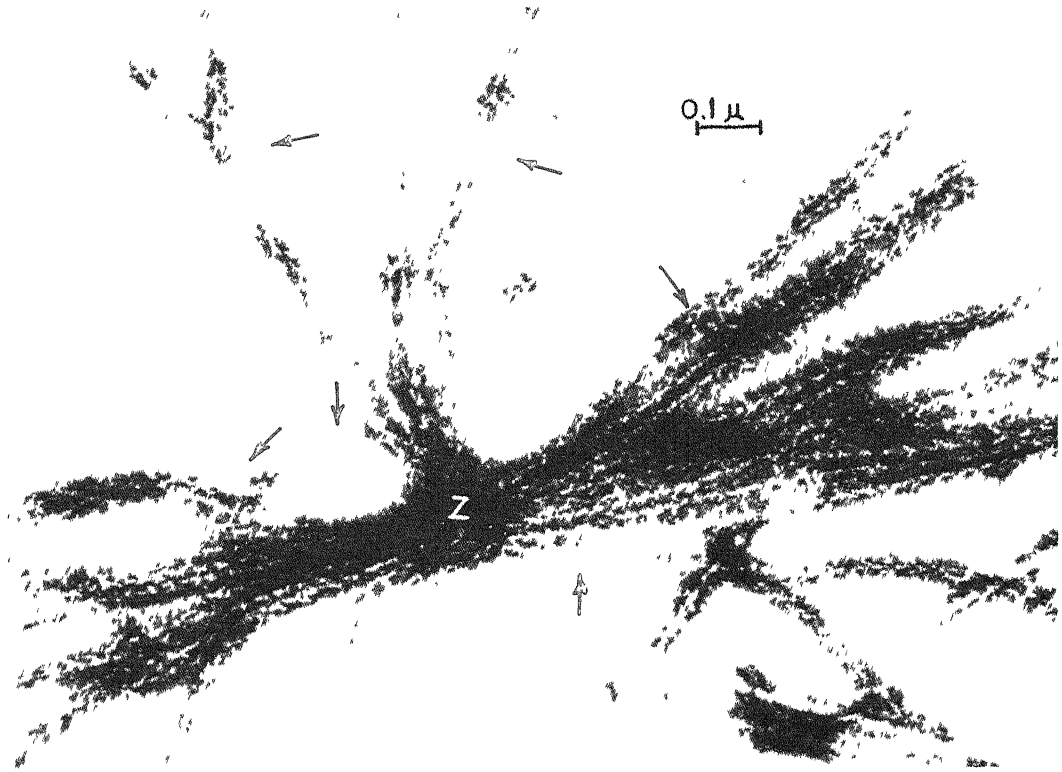


Fig 1 62 An isolated I segment has been treated with heavy meromyosin which attaches to the thin filaments producing arrowhead-like structures (note places indicated by arrows) These are polarized and point outward from either side of the Z line Segment prepared by blending glycerol-extracted muscle in a medium which separates thick and thin filaments (From Huxley 1963)

(Fig 1 64) has been suggested in which the cross-bridge projections are attached to the backbone of the myosin filaments at a distance of 60 Å from the center of the thick filaments and then extended outward beyond the myosin backbone for an additional 70 Å, thus, the total projection from the center of the myosin backbone is 130 Å The distance between the center of the thick filament and the surface of the actin filament is about 190 Å in a muscle at rest length Thus, these observations indicate that, at rest, cross-bridges do not reach the actin filaments It has also been noted that, at rest, the cross-bridges are not precisely fixed in position on the thick filaments

The x-ray diffraction patterns of the actin filaments show a nonintegral double helical structure, with subunits repeating at 54 6 Å intervals along both strands but the position of any subunit on one strand is staggered relative to the position of its nearest neighbor subunit on the other strand by half a subunit period (27 3 Å) The chains twist around each other with cross-over points at 360 to 370 Å intervals so that the pitch of the helix formed by either of the two chains is 720 to 740 Å (Fig 1 65) Off meridional reflections from both actin and myosin filaments indicate that the helices in the filaments may be able to twist and untwist to some extent, but the axial repeat of the subunits remains rather constant

When all the ATP in the muscle is lost and the muscle goes into rigor, it is believed that a major portion if not all of the cross-bridges become attached to actin filaments Since the subunit repeat of the actin and the spacing of the cross-bridges are entirely different, a matching of the cross-bridge and its binding site on the actin monomer occurs only once every several thousand angstrom units along the filaments, therefore, bonding of the cross-bridge to actin requires a change in conformation of the actin and/or cross-bridge molecules in going from the resting to the contractile state. Thus, x-ray diffraction studies of muscle in rigor reveal marked changes in the attach-

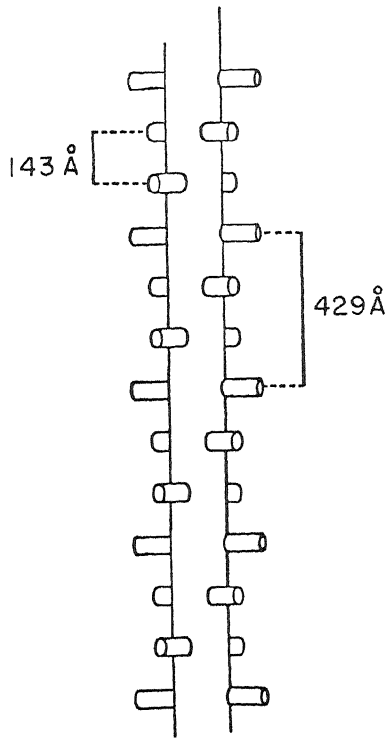
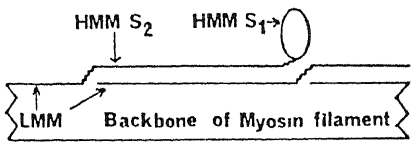


Fig 1 63 Diagram of cross-bridge arrangement on thick myosin-containing filaments of frog sartorius muscle which would account for the observed x-ray pattern (From Huxley 1969)



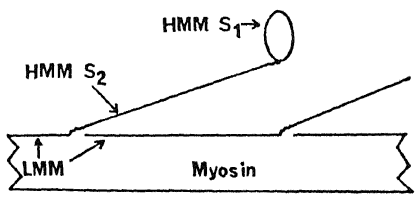


Fig 1 64 Suggested behavior of myosin molecules in the thick filaments The light-meromyosin (*LMM*) part of the molecule is bonded into the backbone of the filament while the linear portion of the heavy-meromyosin (*HMM*) component can tilt further out from the filament (by bending at the HMM-LMM junction) allowing the globular part of HMM (that is the S_1 fragment) to attach to actin over a range of different side-spacings while maintaining the same orientation (From Huxley, 1969)

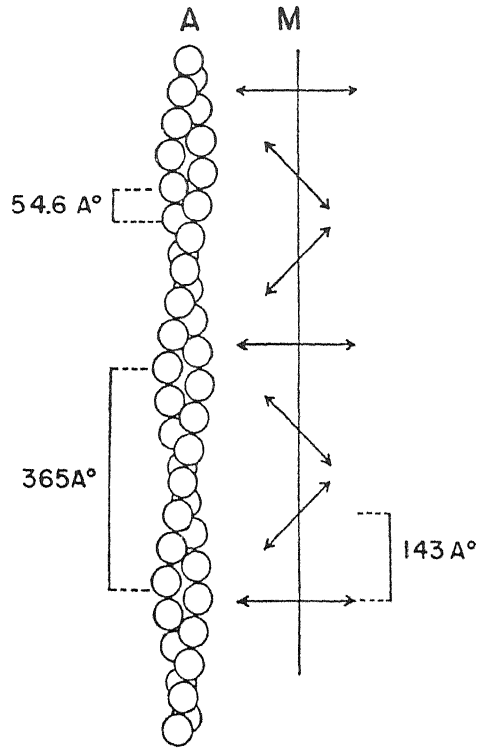


Fig 1 65 Diagram of the arrangement of G actin monomers in actin (*A*) filaments derived from x ray diffraction and electron microscope observations Both the pitch of the helix and the subunit repeat differ from those of the myosin (*M*) filaments indicated schematically alongside Thus cross-bridges between filaments would act asynchronously and a sequence of them would develop a fairly steady force as the filaments moved (From Huxley 1969)

ment of the cross-bridges to actin but no change in the actin filaments per se The 429 Å repeat pattern of the cross-bridges disappears completely indicating that a large portion of the cross-bridge structure is displaced from the position it normally occupies during rest The pattern observed in rigor indicates a helical pitch of 720 Å which approximates the helical pitch of the actin filaments This observation confirms that, in rigor, a high proportion of the cross-bridges attach to the actin filaments.

The axial separation of the cross-bridges remains unchanged at 143 Å during rigor although the bridges rotate away from the myosin backbone, and thus extend more perpendicularly toward the actin filaments. With this reorientation a span of 5 cross-bridges (5 × 143 Å) closely matches the 720 Å pitch of the

actin filaments permitting an increase in the number of bonds formed

The Myosin Filament Model Huxley has suggested that cross-bridges are attached at their bases to the backbone of the myosin filaments by a short subunit tail which is soluble in physiological solutions This tail (approximately 400 Å in length) is attached to the thick filament only at one end This type of attachment makes it possible for the cross-bridges to move away from the backbone toward the actin filament As noted above the attachment of the heavy meromyosin (HMM) tail to the light meromyosin (LMM) of the thick filament backbone is susceptible to trypsin digestion and, therefore, probably represents a less than helical, flexible region of the molecule Similarly, the junction between the linear part of the HMM tail with the globular head is also susceptible to enzymatic digestion (with papain) and this site of attachment probably also constitutes a flexible site Therefore, the presence of two flexible points on the tail of the cross-bridge would give the HMM head the mobility needed for it to become properly oriented towards the actin filament during contraction (Fig 1 66*A*, *B*, and *C*)

On the basis of both mechanical and biochemical considerations, the flexible tail of the cross-bridge cannot itself account for the development of tension during contraction inasmuch as the ATPase activity of HMM remains with the globular head after the linear tail is removed by enzymatic treatment However, the orientation of the linear tail to the thick filament backbone suggests that the tail serves to sustain the tension which is developed by the globular end of the HMM cross-bridge Observations of interactions between cross-bridges and actin filaments indicate that globular HMM binds tightly to actin at least during part of the contraction cycle, therefore, Huxley has suggested that the linkage of the globular HMM to actin is the most stable and rigid (al-

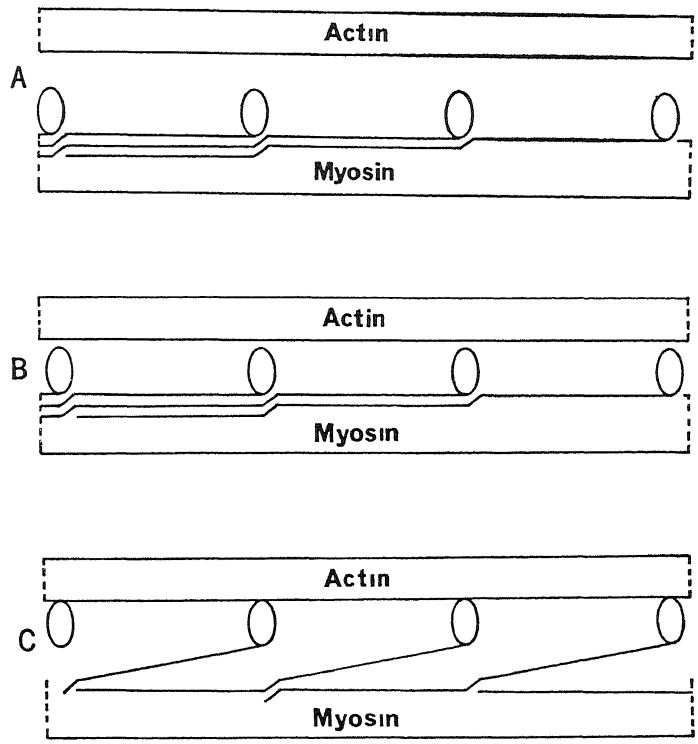


Fig 1 66 Diagram showing relative positions A of filaments and cross-bridges at two different interfilament spacings [(*A*) 250 Å and (*B*) 200 Å] corresponding, in frog sartorius muscle to sarcomere lengths of ~2 0 and ~3 1 μ The x-ray diagram (not shown) suggests that in a relaxed muscle the cross-bridges do not project very far toward the actin filaments During contraction or in rigor the cross-bridges could attach to the actin filaments by bending at two flexible junctions as shown in (*C*) (From Huxley 1969)

belt temporary) attachment between the thick and thin filaments In this hypothesis, tilting of the globular HMM head of the cross-bridge at the site of attachment to the actin would initiate the sliding of the actin filament along the myosin filament (Fig 1 67*A*) In an alternative hypothesis, a similar sliding of the actin filament is initiated by a mechanism involving oppositely directed movements of the paired subunits in each globular head of the cross-bridges (Fig 1 67*B*)

In summary, *1*) the ATPase activity of myosin and the high affinity of the cross-bridges for actin is associated with the globular HMM head of the cross-bridge, and 2) the sliding of the thin (actin) filament along the thick (myosin) filament depends on a characteristic cycle of attachment, movement, and separation of the HMM head of the cross-bridge to successive sites along the actin filament as well as on the adaptability of the long (400 Å) flexible HMM tail of the cross-bridge

Davies (1963) has advanced an hypothesis of muscle contraction based on the molecular structure of the contractile elements This theory has the following assumptions *1*) the ATP-binding site is located at the end of a mobile polypeptide which is a component of the HMM cross-bridge The bound ATP is held by three electrostatic positive charges (probably Mg^{++} ions linked to negative charges on the protein) *2*) There is a fixed negative charge (perhaps phosphorylserine) inside the HMM, near the mobile polypeptide which repels the negatively charged ionized ATP so that during rest the assymmetric polypeptide chain is uncoiled *3*) When the muscle is activated, the depolarization of the sarcoplasmic reticulum liberates enough calcium to make chelate links between the bound ATP of the extended HMM polypeptide and the bound ADP of the F actin This step provides the link for electrostatic bond formation between the actin and myosin filaments and neutralizes the negative charge on the terminal phosphate resulting in the abolition of the repulsive force within the HMM Thus the extended polypeptide recoils and forms an α helix by twisting and forming hydrogen bonds as well as hydrophobic bonds If the magnitude of the force thus created is great enough to overcome the resistance of the series elastic elements, actin molecules will be pulled along the myosin filament and the muscle will shorten *4*) The ATPase activity of HMM is inside the cross-bridge with no access to free cytoplasmic ATP When the polypeptide chain coils, bound ATP is moved within the range of activity of the HMM ATPase, which then hydrolyzes the ATP and breaks the link between

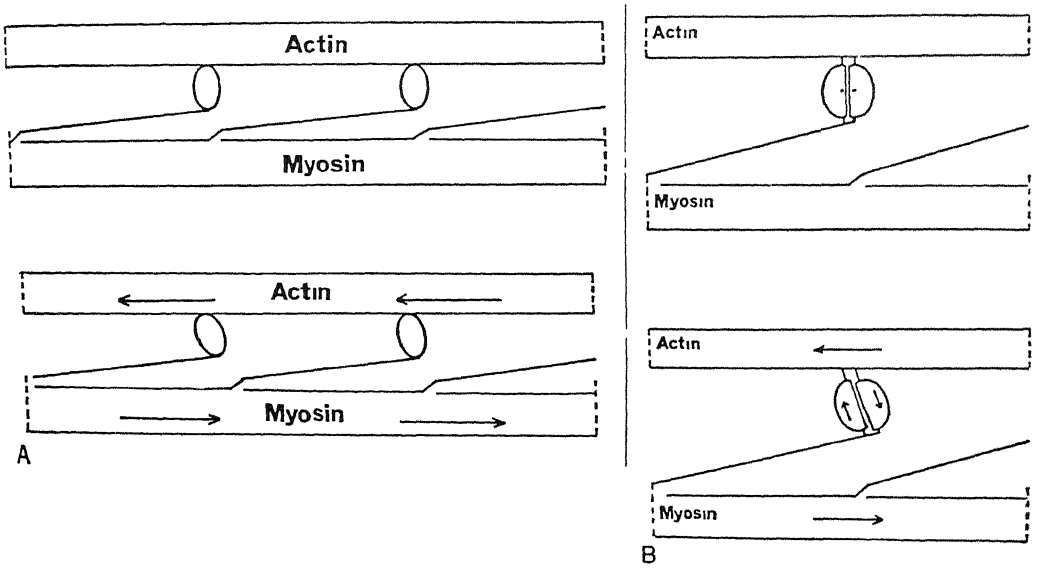


Fig 1 67 Diagram illustrating possible mechanisms for producing relative sliding movement by tilting of cross bridges *A* If separation of filaments is maintained by electrostatic force-balance, tilting must give rise to movement of filaments past each other *B* A small relative movement between two subunits of myosin could give rise to a large change in tilt by the mechanism shown (From Huxley 1969)

actin and myosin ADP bound to the cross-bridge is rephosphorylated in the next step by cytoplasmic creatine phosphate

ADP may also slowly exchange with cytoplasmic ATP The formation of ADP would thus serve to inhibit the activity of ATPase When ATP is regenerated, its negative charge will repulse the negative charge inside HMM so that the latter is extended along the actin filament 5) This process repeats itself as long as Ca^{++} ions are present However, because Ca^{++} ions are quickly sequestered by the sarcoplasmic reticulum, this active state is soon ended The sequestration of Ca^{++}, being an active process, is also accomplished with the energy derived from ATP According to Davies the energy for contraction is provided by the formation of hydrogen bonds which had been previously split by the action of ATP

On the basis of the above theory, Davies has suggested a number of explanations for many of the phenomena of muscle contraction, i e, *1*) the number of Ca^{++} ions required for maximal contraction must be equal to the number of cross-bridges, *2*) the length of latent period is equal to the time it takes the released Ca^{++} ions to reach the sites of linkage between actin and myosin (the latent period also includes the time taken up by the events leading up to the release of Ca^{++} and by the actual activation of actomyosin to contract), *3*) following stimulation, the early volume alteration of muscle (first increase and then decrease) and the phenomenon of latency relaxation may be associated with differences in electrostriction between the states of Ca^{++} tightly bound in the sarcoplasmic reticulum, in free solution in the sarcoplasm, and in the postulated bond between the nucleotides, *4*) the increase in the resistance to stretch following stimulation is associated with Ca^{++} chelate links which form prior to contraction, *5*) activation heat may be due to the dissociation of Ca^{++} from their tightly bound sites in the sarcoplasmic reticulum

The Single Muscle Twitch

MECHANICAL CHARACTERISTICS OF MUSCLE

A relaxed, unstimulated muscle is elastic and offers little resistance to stretch The force needed to stretch the muscle increases, however, as the muscle is being extended The tension developed in the muscle as a consequence of stretching does not increase linearly but approximately exponentially as the length of the muscle increases (Fig 1 68)

If resting muscle is exposed to a small load its elasticity is "rubber-like," that is, it shortens when warmed and lengthens when cooled At greater load, i e at greater initial length, the elasticity is normal and not "rubber-like "

The elasticity of the resting muscle has been ascribed to properties of the contractile material itself or to the sarcolemma which involves the plasma membrane of the muscle fiber, the "basement membrane" material coating the outer surface of the plasma membrane and connective tissue elements associated with the muscle fiber surface At moderate degrees of stretch it is unlikely that the sarcolemma would contribute substantially to the elasticity of the muscle

When a muscle or a single muscle fiber is stimulated by a single stimulus while extended by a moderate load, there is a brief lag between the arrival of the stimulus and the initiation of tension development This lag is referred to as the *latent period* (2 to 3 msec in the frog sartorius muscle) When the tension has reached a value exceeding the stretching force exerted by the load, the muscle shortens rapidly This development of contractile force is transient, and as the contractile force decreases, the muscle lengthens and returns to its relaxed condition (Fig 1 69)

If a muscle is allowed to shorten freely under a given load (isotonic contraction) and this experiment is repeated applying different

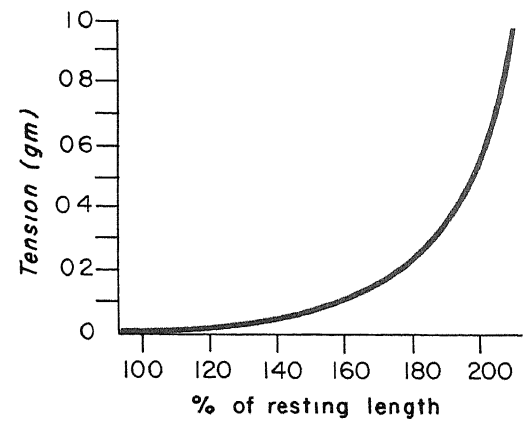


Fig 1 68 Static length-tension diagram of isolated single muscle fiber at rest (From Buchthal et al, 1951)

loads, the degree of shortening as well as the velocity of shortening are found to be dependent on the load With increasing load, the amount of shortening and the velocity of shortening decrease and the latency period increases

If the stimulus is repeated at a sufficiently high rate the muscle will not relax between each stimulation but rather will remain in a contracted state which is referred to as *tetanus* The plateau of such a tetanic contraction exceeds the peak of a single twitch and, in fact, represents the maximum contractility of which the muscle is capable This plateau is maintained until the muscle begins to fatigue at which time it relaxes gradually However, if stimulation is discontinued prior to the onset of fatigue the tetanized muscle relaxes immediately (Fig 1 70)

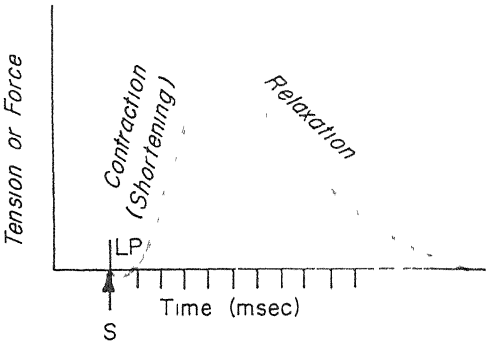


Fig 1 69 Simple muscle twitch *S* stimulus *LP* latent period

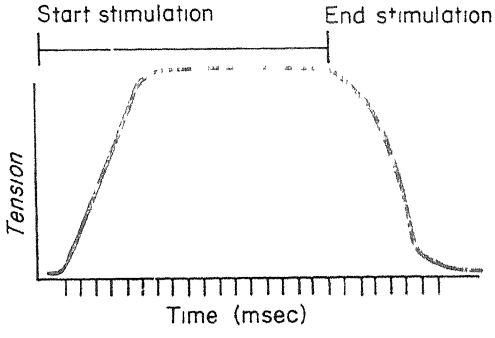


Fig 1 70 Tetanic tension in response to high frequency stimulation skeletal muscle develops and sustains a level of tension greater than that of the simple twitch

CHANGES IN THE EXCITABILITY OF THE MUSCLE DURING THE CONTRACTION-RELAXATION CYCLE

As noted above, a simple muscle twitch is characterized by a short period during which tension is being developed, the contraction phase, followed by a relaxation phase during which the muscle returns to a resting condition If an initial single effective (threshold or suprathreshold) stimulus is followed by a second stimulus of the same or greater intensity, the muscle will or will not respond to the latter stimulus depending on the time interval between the first and second stimulus In other words, there is a brief period following stimulation during which a second stimulus cannot elicit a response (Fig 1.71) This period of inexcitability is referred to as the absolute refractory period (ARP) or as the effective refractory period (ERP) The duration of this period of inexcitability varies from 5 to 50 msec in various skeletal muscles Then the excitability of the muscle returns gradually so that stimuli of greater than threshold intensity will evoke a response although threshold stimuli are ineffective This period is called the "relative refractory period" (RRP)

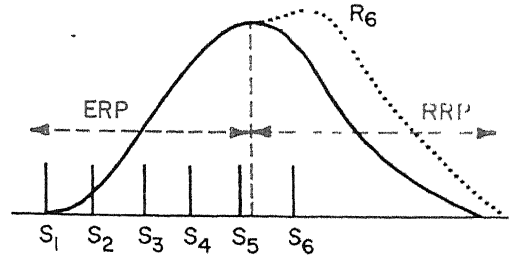


Fig 1 71 Effective and relative refractory periods *S*, stimuli given to the muscle at intervals *R* response of the muscle (a twitch), R_6 shows response to stimulus S_6 *ERP*, effective refractory period *RRP* relative refractory period

The duration of the effective and relative refractory period varies from muscle to muscle (Fig 1 72) In fast muscles, the ERP and RRP are much shorter, therefore, fast muscles can respond to higher frequencies of stimulation For example, the insect flight muscle, which has a total response time (i e , the time required to complete one full cycle of contraction and relaxation) of 2 to 3 msec , can twitch so rapidly that it literally flutters (at a frequency of approximately 300/min)

THE MOTOR UNIT

In normal skeletal muscles, fibers probably never contract as isolated individuals. Instead several of them contract at almost the same

moment, all being supplied by branches of the axon of one spinal motor neuron In other words, when the motor neuron is excited by a stimulus at or above threshold, all muscle fibers innervated by the neuron will be activated The single motor neuron and the group of muscle fibers which it innervates is called a *motor unit* (Fig 1 73), this is the smallest part of the muscle that can be made to contract independently The number of muscle fibers in this unit varies in different muscles, from 2 to 3 to more than 1000, the size of the unit is correlated with the precision with which the tension developed by the muscle is graded (see below)

The gradation of the tension developed by a muscle depends on variation in the number of motor units activated as well as on the frequency at which each unit is stimulated

OTHER CHARACTERISTICS OF A SINGLE TWITCH

All-or-None Response Stimulation at threshold intensity (minimum effective stimulation) evokes a minimal twitch tension There is one response to one stimulus If the intensity of the stimulus is increased stepwise the amplitude of the twitch tension will likewise increase However, with continued increase in the intensity of stimulation, a maximal contractile response is attained after which further stimulation no longer results in a change in the amplitude of the twitch tension (Fig 1 74)

The explanation for these responses lies in the fact that with low intensity stimuli, only the most excitable nerve fibers respond and activate their respective motor units, but as the intensity of the stimuli is increased, more motor units are recruited until finally all motor units are responding at which time the twitch tension is maximal and, therefore, an increase in the intensity of the stimuli can no longer elicit an increased response It is to be noted that the *individual* motor unit responds in the classical "all-or-none" manner Recently, it has been possible to tease single muscle fibers and record their tension, in such preparations the "all-or-none" law is clearly demonstrated

The Series Elastic Element. If a muscle is rapidly stretched during the early part of a contraction, or even during the latter part of the latency period, it can be demonstrated that it is capable of developing a considerably greater tension than that recorded under normal conditions (i e , without stretching) at the same time after stimulation This observation has been interpreted to reveal the presence in the muscle of an elastic element coupled in series with the contractile elements Following stimulation, the shortening of muscle is delayed because the

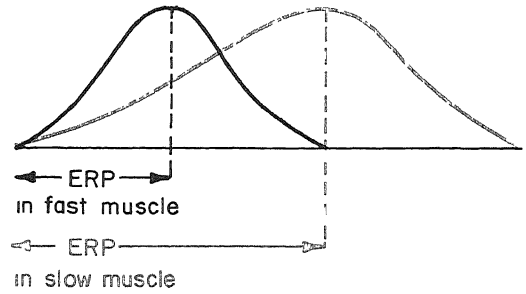


Fig 1 72 Relative duration of effective refractory period (*ERP*) in fast and slow muscles

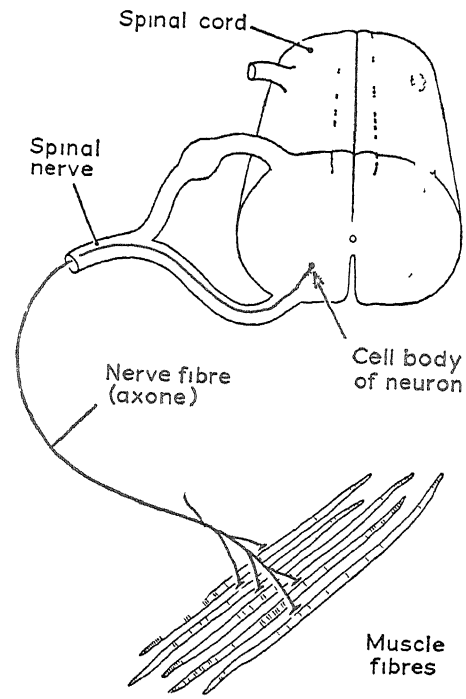


Fig 1 73 Diagram of a motor unit

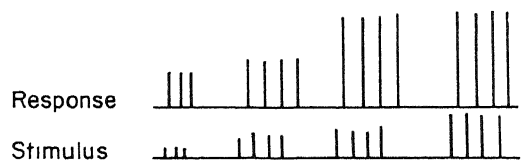


Fig 1 74 Recruitment of motor units with increasing stimulus intensity The response of the individual unit is 'all-or-none but the threshold of excitability varies among the units

contractile elements must first stretch the series elastic elements before a tension can be recorded externally

The quick-stretching experiments also show that the contractile machinery is activated already during the latency period and that this activation, the "active state," very rapidly reaches a maximum which is maintained at a constant level for only a short period of time

The existence of a series elastic element can be demonstrated in an obvious way by quickly releasing one point of fixation of an isometrically contracting muscle and allowing it to shorten to a new equilibrium length The tension then falls to zero before a new level of tension develops which is characteristic for the newly established length This instantaneous fall in tension has been ascribed to the undamped series elastic element

It is likely that at least part (about half) of the series elastic element resides in the contractile system itself and more specifically in the filamentous structure in which the main muscle proteins are arranged, although the suggestion has also been raised that the Z line structure is the site of the series elastic element

Resting Tension. It has been observed that muscle which is neither stimulated nor stretched beyond the normal body length still maintains some degree of tension, a phenomenon which is revealed by the simple fact that if a tendon is cut in vivo its associated muscle shortens Recently, Hill (1970) has adduced evidence that resting tension is due to residual long-lasting interactions between actin and myosin filaments and that it increases with rising ambient temperature

Latency Relaxation. When a loaded muscle is stimulated, the first mechanical response which appears during the latent period is a slight lengthening which is referred to as latency relaxation This phenomenon, which must be due at least in part to the release of resting tension, begins about 1 0 to 1 5 msec after stimulation (frog sartorius) and lasts for about 2 msec

It is likely that latency relaxation reflects the disruption of long-lasting cross-bridge bonding between myosin and actin filaments just prior to the initiation of the more rapid "make and break" cross-bridge bonding associated with the sliding process

Latency relaxation is most marked at longer lengths of muscle and is not at all evident when the muscle length is substantially (10% or more) below the normal body length

ISOMETRIC CONTRACTION

When the two ends of a muscle are held at fixed points, stimulation causes the development of force without change in muscle length (isometric tension) If adequate stimulation is delivered so that all muscle fibers are synchronously contracting, the isometric tension following a single stimulus changes as shown in Figure 1 69 If multiple stimuli of increasing frequency are delivered, the isometric tension follows the same curve as the single twitch however, tension increases until it reaches a maximum level which is maintained until the stimulation is discontinued This sustained maximal isometric tension, as in the case of sustained isotonic contraction is referred to as *tetanus*, and constitutes the maximum tension that the muscle can develop Tetanic tension will decline shortly after the stimuli have ceased and the muscle then relaxes with a characteristic time course Isometric tension is also dependent on the length at which the muscle is kept during the experiment The length tension relationship for isometric contraction (Fig 1 75 *A* and *B*) shows a maximum tension at a length slightly less than the length of the muscle in the body This amounts to about 3 5 kg /cm ^{2} It is important to recognize that the curve 3 in Figure 1 75 *B* is a composite of active and resting tension If the component of resting tension (curve 1, Fig 1 75 *B*) is subtracted, the curve of residual active tension vs time will rise and fall (curve 2, Fig 1 75 *B*) indicating that active tension declines as the muscle is stretched beyond normal body length These length-tension relationships can be explained in terms of the sliding filament theory of muscle contraction When tension is minimal, the length of the muscle is referred to as initial length (L_i) With the muscle length at L_i (considerably less than body length) the actin filaments overlap each other in the center of the sarcomere (Fig 1 56 *D*) to such an extent as to reduce the number of active sites available for interaction with myosin As the sarcomere is stretched, the "abnormal" overlap of the actin filaments is progressively reduced and accordingly, more active sites become available for linkage to the myosin cross-bridges. As the

muscle approaches the normal body length, the length at which tension becomes maximal is defined as optimal length (L_o) In this condition the position of the actin and myosin filaments is such as to provide a maximum number of active sites for interaction (Fig 1 56 *B*) As the muscle is further stretched beyond the optimal length the overlap of actin and myosin filaments is again progressively reduced and contractile tension diminishes until a point is reached at which there is no overlap (Fig 1 56 *C*) and, consequently, no active tension can be developed At this point the muscle is at maximal length (L_{max})

In isometric contraction the shortening of any contractile element is minute and a moderate configurational change associated with the lateral projections (cross-bridges) of the myosin filaments can account for the development of tension This would imply that, during isometric contraction, a large number of subunits in the myosin filaments could interact synchronously to develop tension, the number depending on the degree of overlap The highest isometric tension has been recorded when the zone of overlap extends the entire length of the myosin filaments over which cross-bridges are present This implies that the development of maximum tension during isometric contraction is related to the condition allowing maximal interaction between myosin and actin During isometric contraction, how-

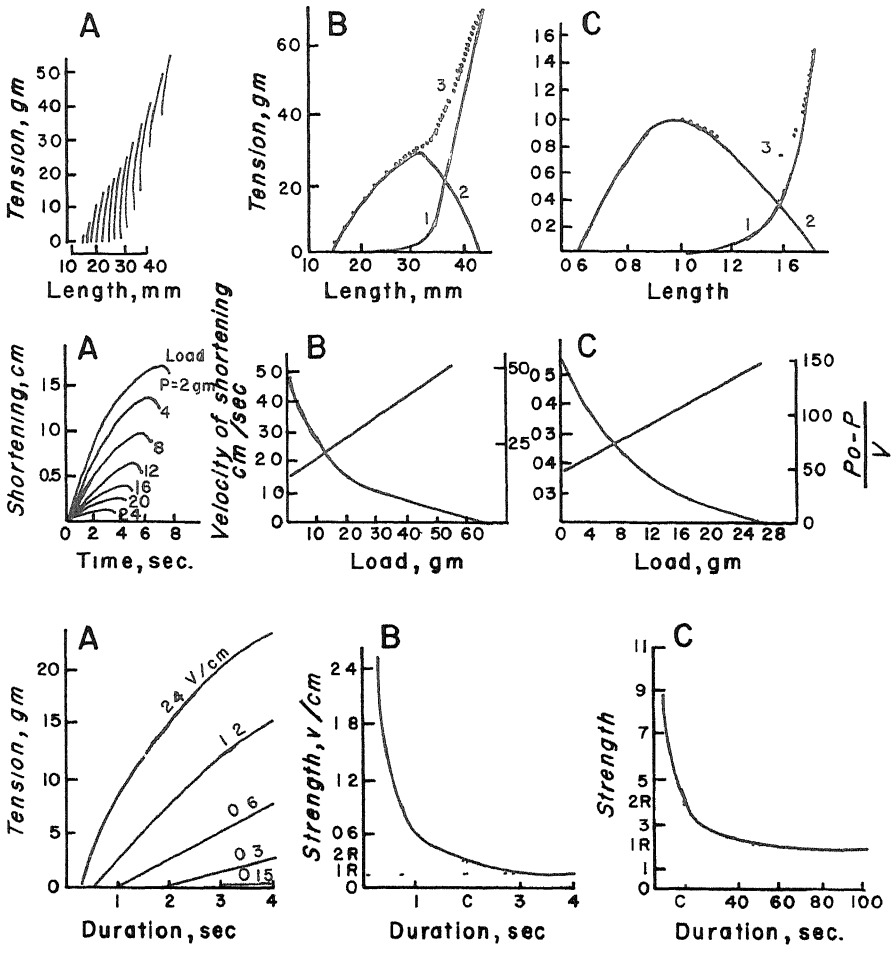


Fig 1 75 The length-tension load-shortening velocity and strength-duration relations of the frog sartorius and the rabbit uterus The first picture in each row is an original record from an experiment with the uterus the second is a plot from the first record to be compared with the third in each row illustrating the same experiment with the frog sartorius Note the general similarities between the frog sartorius and the rabbit uterus in each of the three relations (Csapo 1960)

ever, the conditions for interaction between myosin and actin can be restricted, since it is mainly the subunits located at the ends of the myosin filaments which are involved, and since the length of the sarcomere determines the number of actin-myosin interactions

ISOTONIC CONTRACTION

When one end of the muscle is free, it shortens while exerting a constant force The curve of isotonic contraction begins with a short period of increased rate of shortening which is followed by a "linear period," i e , a period of shortening at a constant rate during which the force of muscular contraction just overcomes the gravitational force The degree of shortening of a muscle bears a direct relationship to the load upon which it exerts force A progressive increase in the frequency of stimulation (at suprathreshold levels) will evoke a partial and then a complete state of tetanus which, under conditions of isotonic contraction, is characterized by sustained maximal shortening (Fig 1 70) The additional shortening observed during tetanus as compared with the shortening observed during a single twitch indicates that in the latter circumstance the muscle is not fully activated

FORCE-VELOCITY RELATIONSHIP

If a muscle is contracting isotonically (i e , shortening against a load), the velocity of contraction is inversely related to the load which the muscle is lifting If the muscle is unloaded it shortens with maximum velocity and as the load increases, the velocity of shortening declines When the load reaches a maximum (which is equal to the maximum isometric tension the muscle can develop), the muscle no longer shortens The relationship between load and the velocity of shortening is hyperbolic except at the extremes as shown in Figure 1.76 Hill (1970) and co-workers adduced the following empirical equation on the basis of data obtained from frog sartorius muscle tested at 0° C

$$(P + a)(V + b) = \text{constant} \qquad (1)$$

where P is the load applied to the muscle, V is the velocity of shortening and a and b are constants with dimensions of force and velocity, respectively It can be seen from Figure 1 76 that when $P = P_o$, $V = 0$, P_o being defined as

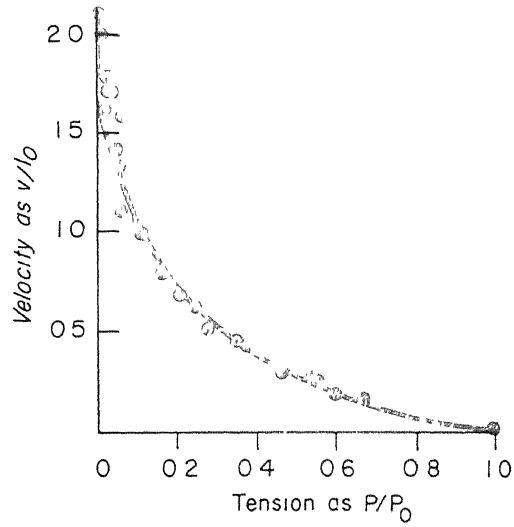


Fig 1 76 Force velocity relation, derived from experiments made with velocity as the independent variable Velocity as v/l_o, force as P/P_o Twenty four measurements in three experiments are shown as circles The curve was drawn from the usual equation with constants given in the text (From Hill 1970)

the maximum isometric tension that the muscle can develop, therefore

$$(P_o + a)b \quad \text{constant} \qquad (2)$$

Therefore

$$(P + a)(V + b) \quad (P_o + a)b \qquad (3)$$

Equation 3 can be rearranged to take the following form

$$(P + a)V \quad (P_o \quad P)b$$

This equation states that the rate of extra energy liberation during contraction (over and above the energy used for activation and maintenance of tension) is inversely and linearly proportional to load This follows because both P and a have dimensions of force and $V \quad dx/dt$, the rate of change in length. Thus, $(P + a)V$ is the rate of release of energy (force x distance) with time The right hand term $(P_o - P)b$ increases as load (P) decreases, b being a constant of proportionality The significance of this observation was recognized long before the advent of sliding-filament theory by Hill (1938) who postulated "that the inverse dependence of energy release upon load could be explained if each active point in the molecular machinery (i e , those sites which generate tension and effect shortening) could exist in one of the two following states *1)* in combi-

nation with each other and thereby developing tension or 2) uncombined or free, and thereby able to take part in chemical transformations At any instant the distribution of the active points between those two states would be determined by the load which the muscle bears If the ends of the muscle are fixed, so that no external shortening occurs and the tension developed is at its maximum (isometric contraction), all active points would tend to be in the combined state Thus, P_o (maximal tension, P at zero velocity) should be determined by the total *number* of active points With lighter loads fewer points would be in the "cimbined" state and more in the "free" state, until both shortening velocity and the rate of energy release are maximal when the muscle shortens freely at zero load Here, because essentially all active points are in the free state, the maximal shortening velocity (V_{max}) will be determined by the intrinsic rate at which the active points are able to undergo cyclic position changes which effect shortening As a result, the limiting *rate* at which the active points of the contractile machinery couple and uncouple should determine both V_{max} and the maximal rate of energy release"

If the term "active point" is replaced by the phrase "actin-myosin interaction", Hill's entire analysis can be translated into present day concepts of the contractile process including the correlation of V_{max} and the rate of myosin ATPase activity (Bárány, 1967)

Hill observed that the constant, a, was independent of temperature and had a value of 15 g when the maximum isometric tension (P_o) was 60 g Therefore, he postulated that in general the value of the fraction, a/P_o, was equal to 0 25

From (3) it follows that when

$$P = 0 \text{ and } V = V_{max}, \text{ then}$$

$$a\ V_{max} = b\ P_o$$

and, rearranging,

$$V_{max} = b/a\ P_o \qquad (4)$$

Thus, substituting Hill's values (P_o = 60 g , a = 15 g)

$$V_{max} = 4b \qquad (5)$$

Unlike a, b is temperature dependent exhibiting a Q_{10} ranging from 2 to 2 5 Thus, b appears to be related to chemical processes underlying the mechanism of shortening The value of b varies from muscle to muscle

Further experiments involving the velocity of contraction, load, heat production and work performed by muscle provided independent data supporting the validity of the constants, a and b, as employed in the Hill equation In these experiments the apparatus used to measure the generation of heat by muscle consisted of an elaborate system of thermocouples (referred to as a thermopile) connected to an ultrasensitive galvanometer (Fig 1 77)

HEAT PRODUCTION DURING THE CONTRACTION-RELAXATION CYCLE

Muscle contraction is associated with heat production, a sign of the chemical and physical events associated with contraction A rise of temperature in muscle can be measured to within about 10^{-5} ° C, and the time-course of heat liberation can be resolved within a few milliseconds with the thermoelectric method, a technique which provides the only means to follow the rapid internal physical and chemical changes occurring during and after a contraction

The heat production of unstimulated muscle amounts to 0 0002 cal /g min and is referred to as "resting heat " The heat produced during a single muscle twitch is about 0 003 cal /g and appears as *1*) an initial very rapid heat production which starts before any shortening has occurred, and *2*) heat liberated during shortening in proportion to the amount of shortening of the muscle If the muscle performs

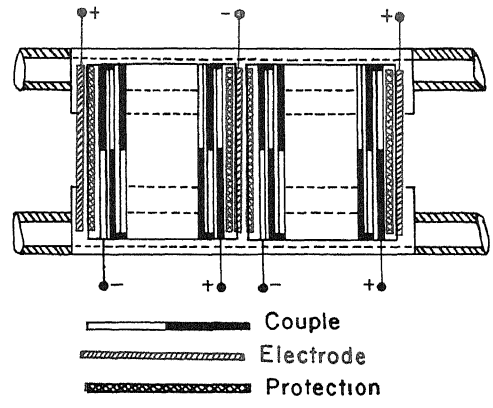


Fig 1 77 The pile consists of two identical frames supporting three thermocouples connected with indifferent electrodes The couples are isolated by a thin sheet of mica (Aubert 1956)

work, additional heat is liberated in proportion to the work done (see Fenn effect below) The initial heat liberation has been associated with the development of the active state, it is independent of the extent of shortening of load, of the velocity of shortening, of temperature of oxygen supply, and of work done—and is referred to as "activation heat "

In the second stage of heat liberation the amount of heat released during shortening (Figs 1 78 and 1 79)—referred to as the "shortening heat"—is proportional to the degree of shortening

$$H_s \quad \alpha x \tag{6}$$

where α is the coefficient of the heat of shortening and x is the distance representing the decrease in length of the shorteed muscle

In evaluating the total energy expenditure of muscle, one must consider the mechanical work accomplished as well as the heat produced Thus, the rate of release of energy during shortening is

$$(d/dt)\, E_s = (d/dt)\, (H_s + W) \tag{7}$$

where E_s is the total energy expenditure involved in the overall shortening process, H_s is the heat of shortening and W is the work done on shortening W equals Px where P is the load applied to the muscle and x, as defined above, is the amount of shortening Substituting for heat and work in equation (7) we have

$$(d/dt)\, E_s = (d/dt)\, (\alpha x + Px) \tag{8}$$

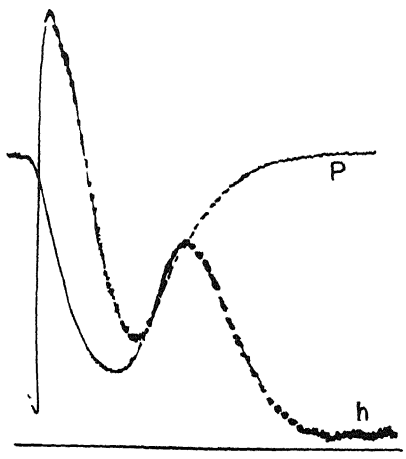


Fig 1 78 Simultaneous oscilloscopic recording of mechanical tension (*P*) and heat production (*h*) during an isometric twitch (Aubert 1956) Note that in this recording increasing tension corresponds to a downward deflection

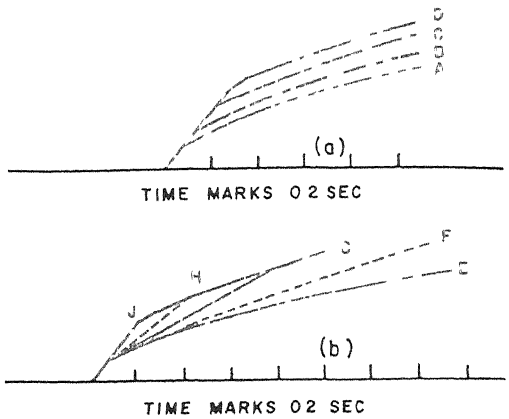


Fig 1 79 Heat production during isotonic shortening (tetanus 0° C) (*a*) shortening different distances under constant load of 1 9 g (*A* 0 0 mm isometric) (*B*) 3 4 mm (*C*) 6 5 mm (*D*) 9 6 mm (*b*) shortening constant distance 6 5 mm under various loads (*E*) isometric (*F*) 31 9 g (*G*) 23 7 g (*H*) 13 8 g (*J*) 1 9 g The ordinate represents total heat produced (Hill 1938)

When P is constant

$$(d/dt)\, E_s \quad [dx/dt]\, (\alpha \quad P) \tag{9}$$

But dx/dt equals V, the velocity of contraction, therefore

$$(d/dt)\, E_s \quad V\, (\alpha + P) \tag{10}$$

When energy expenditure was measured at loads ranging from 0 to P_o, $(d/dt)\, E_s$ was found to be inversely proportional to load as follows

$$(d/dt)\, E_s \quad \beta\, (P_o \quad P) \tag{11}$$

The constants a and b of equations (1) to (3) were estimated by analysis of the relationship between load and the velocity of shortening, whereas, the constants α and β of equations (10) and (11) were calculated from heat production measurements The value for a proved to be 0 25 P_o, whereas, the value for α varied with load as follows $\alpha = 0\ 18\ P_o + 0\ 16\ P$, thus, α proved to be dependent on changes in P which may range between 0 and P_o Accordingly α has a lower limit of 0 18 P_o (when P 0) and an upper limit of 0.34 P_o (when P P_o) The constant, a, and α both have the dimension of force

The Fenn Effect. Fenn (1923, 1924) observed that the total energy released by a muscle (work + heat) increases as muscle work increases This fundamental observation revolutionized the concept of muscular contraction because, previously, it had been believed that when a muscle contracts a preset amount of

additional energy is fed into the contractile machinery (in association with the assumption by the muscle of new elastic characteristics), and that this present increment of energy could be proportioned between work and heat, the total amount of energy (work + heat) remaining constant irrespective of the amount of work performed by the muscle In other words, it was believed that when the muscle assumes new properties associated with the contractile state, the total energy made available to it for release as work and/or heat was determined by the activation process Fenn's finding indicated that the energy release is determined not only by the activation process but also by the work load imposed on the muscle Thus the mechanical function was shown to control chemical reactions in muscle cells, the molecular mechanism being finely attuned to the work demand, (i e , the performance of a given increment of muscular work calls for the generation and delivery of a commensurate increment of energy)

There is no heat liberated during relaxation except when a load stretches muscle In this case the liberated heat corresponds to the work done by the muscle during shortening, this work is degraded into heat as the muscle is stretched

Further heat is liberated during recovery This "recovery heat" is associated with oxidative chemical synthetic processes The recovery heat is about equal to the total of the initial energy consisting of activation heat, shortening heat and work Since about half this initial energy is derived from chemical reactions that are reversed during recovery it appears that about half the energy associated with the recovery reactions is wasted as heat

From a knowledge of the energetics of muscle contraction it is possible to estimate the mechanical efficiency of a muscle, that is, the fraction of the chemical energy liberated during contraction and recovery that is turned into mechanical work This efficiency is generally not higher than 20 to 25%, part of the energy being wasted during contraction and part during recovery

THE ACTIVE STATE

The active state of muscle which develops following an adequate stimulus commences with the onset of latency relaxation The intensity of the active state (I_{as}) is defined as the "maximal isometric tension which contractile components can develop " Only a part of this maximal tension is actually manifest in a twitch because of the damping effect of series elastic elements, i e , the "spring-like" quality of the series elastic elements initially absorbs the energy developed during the onset of the active state as in the sarcomere

I_{as} can be determined at various times after stimulation by measuring the maximum load that a muscle can sustain after a small quick stretch Shortly after a muscle starts to contract (about 20 msec after stimulation for frog muscle), a quick stretch demonstrates that, in fact, the load-bearing capacity of the contractile element has reached its peak Thereafter I_{as} declines exponentially and by the time the muscle has developed peak tension, I_{as} has subsided substantially

I_{as} is affected by the nature and concentration of ions in the medium in which the muscle is immersed and by drugs such as epinephrine, caffeine, nitrate, and quinine The duration of the active state is inversely related to temperature ($Q_{10} \simeq 2$)

Gasser and Hill (1924) had observed that the properties of frog sartorius muscle change drastically soon after it is stimulated For instance, stimulated muscle develops a resistance to stretch which is not observed at rest Furthermore, as noted above, if the muscle is briefly stretched shortly after stimulation it develops maximum tension, however, if the muscle is briefly stretched at progressively lengthening intervals following stimulation it develops increasingly lesser levels of tension Conversely, if a muscle is stretched prior to stimulation and then released to the resting length at various intervals after the stimulation, it can be observed that quick release causes an initial loss of tension which is followed by a secondary development of tension However, the amplitude of this redeveloped tension is dependent on the exact time interval after stimulation that the muscle is subjected to the quick release maneuver These early observations provided the basis for the formulation of the concept of the "active state" by Hill (1927) Because maximal tension (in a quick stretch experiment) is always observed almost immediately after stimulation, Hill postulated that in the simple twitch response all the contractile elements undergo a rapid transition from a fully resting state to a fully activated state which persists

for a variable period of time, and then declines before peak contractile tension is reached (Fig 1 80)

The quick stretch and quick release experiments of Hill have been confirmed and extended by several other investigators Initially the results of these experiments were explained in terms of "viscous and elastic" components of muscle However, with the advent of electron microscopy and the new insights gained into the subcellular structure of the contractile elements, less significance has been ascribed to the role of the "viscous" elements of muscle in the contraction process It also has become apparent that the instrumentation used initially to map the contour of the active state was rather insensitive Therefore, Hill (1970) has repeated his early experiments and revised his concept of the active state on the basis of the sliding filament theory of muscle contraction, as well as the newer data obtained with improved recording techniques For example, the early data suggested that the active state attained maximum intensity instantaneously and then remained at a plateau for 10 msec or longer It is now known that the active state requires approximately one-third to one-half of the contraction time (i e , time elapsing between initiation of tension development and attainment of peak tension) to reach maximal intensity and that it is not maintained at a plateau but instead declines immediately with a characteristic time dependence

Using repetitive stimuli the contour of the active state can be mapped by measuring the distance between the peak and trough of consecutive twitches With this technique it has been observed that the amplitude (intensity) of the active state is dependent on the resting length of the muscle, that maximal contractile

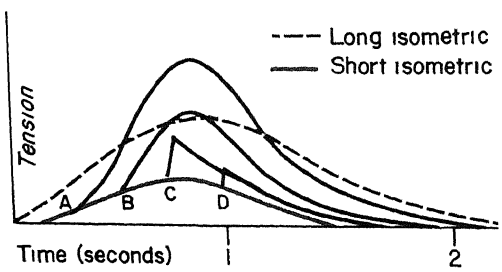


Fig 1 80 Diagram of the effect on the tension developed of stretching a muscle at various instants (*A B, C D*) during a twitch copied from records (From Hill, 1970)

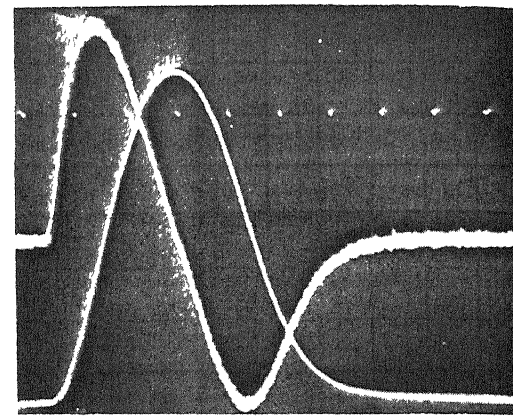


Fig 1 81 A technique for mapping the contour of the active state The monophasic curve is a recording of isometric tension and the biphasic curve is a recording of *dp/dt* the rate of change in the development and release of tension (From the records of one of the authors Penefsky)

tension is obtained at a sarcomere length of 2 4 to 2 5 μ and that various agents which alter the contour of the active state may also alter the contour of the action potential

The quick stretch and quick release methods for mapping the time course of the active state interrupt the instantaneous changes and provide only discontinuous "steady state" glimpses of the total time course Therefore, a new technique has been developed which does not require any manipulation of the muscle and provides a continuous map of the active state contour In this technique, tension is transduced into an electrical signal which, in turn, is differentiated so that the change in muscle tension as a function of time, dP/dt, can be recorded for a muscle twitching isometrically (Fig. 1 81) Inasmuch as the stretching of the elastic elements (which develops the isometric tension) is a reflection of the state of activity of the contractile elements, the dP/dt curve in effect indicates the instantaneous changes in the contour of the active state Observations recorded by this "continuous" technique have confirmed *1*) that the maximum intensity of the active state is not developed instantaneously but requires a measurable period of time, *2*) that after attaining maximum intensity, the active state declines immediately, *3*) that the intensity of the active state depends on the initial length of the muscle prior to stimulation, and *4*) that various conditions which increase contractile tension affect either the peak or the duration of the active state.

Activity of the Contractile Mechanism

STIMULATION VIA MOTOR NERVE

The activity of muscle is controlled by the central nervous system through the motor innervation of the myofibers Each motor nerve fiber splits up into a number of branches that make contact with the surface of the individual muscle fibers with several bulb-shaped endings These endings are arranged in a group, and with a specialized structure of the surface of the muscle fiber they form a motor end-plate

The stimulus conducted along the nerve fiber affects the nerve ending to release acetylcholine, which diffuses across the minute gap separating the nerve ending from the muscle fiber plasma membrane At the motor end-plate, this membrane shows a complex series of invaginations forming minute crypts At this area of the muscle fiber surface, there is a high concentration of acetylcholine esterase which splits (hydrolyzes) the acetylcholine set free during stimulation of the nerve ending

At rest, the electrical potential difference across the muscle cell membrane (i e , the *resting potential*) amounts to approximately −90 mv The release of acetylcholine affects the ionic permeability of the plasma membrane of the muscle fiber at the neuromuscular junction (postjunctional membrane) Due to the difference in concentration of Na^+ ions (high outside) and K^+ ions (high inside), the increased ion permeability gives rise to a sudden influx of Na^+ ions and an efflux of K^+ ions through the plasma membrane This short-circuits the adjacent parts of the plasma membrane and causes a drop in the membrane potential to the level where the membrane becomes electrically excited (i e , the membrane potential rises to the critical firing level, approximately −50 mv) and a wave of depolarization spreads along the muscle fiber (in the same manner as the nerve impulse is propagated along an axon) The electrical response developed as a consequence of the acetylcholine effect on membrane permeability can be recorded as an action potential of approximately 130 mv with the inside of the membrane briefly becoming positive The resting potential of about −90 mv is, therefore, not only abolished but, as in the nerve cell, the distribution of positive and negative charges is reversed The conduction of one impulse involves only a minute amount of Na^+ ions (about 4×10^{-12} moles/cm ^{2}) entering the muscle fiber and an equivalent leakage of K^+ ions to the extracellular space Through a recovery mechanism, involving active transport processes, the ion concentration differences between the muscle cell interior and the extracellular space is re-established by Na^+ extrusion and K^+ uptake

The activity of acetylcholinesterase rapidly abolishes the effect of acetylcholine The electrical response of the plasma membrane of the muscle fiber is therefore efficiently controlled by the nerve impulses reaching the motor end-plate Even without any such impulses there are minute spike potentials developed at the motor end-plate region of the muscle plasma membrane (*end-plate potentials* or *epp*) which have been ascribed to a spontaneous setting-free of minute "packages" of acetylcholine The concentration of acetylcholine that results from this "leakage" is, however, not high enough to produce a sufficient permeability change in the plasma membrane of the muscle fiber to raise the membrane potential to −50 mv , which as noted above, is the threshold for the development of a propagated electric response The minute spike potentials therefore remain localized to the end-plate region

PROPAGATION OF EXCITATION ACROSS AND WITHIN A MUSCLE FIBER

The impulse conducted along the muscle fiber from the end-plate region accounts for a spreading of a state of excitation *along* the fiber The activation of the contractile elements which are distributed over the whole cross-section of the fiber depends on spreading of the excitatory state from the plasma membrane transversally *across* the muscle fiber

The action potential is associated with currents flowing through the interior of the fiber These currents, however, are not directly responsible for activating the contractile mechanism It is, furthermore, not possible for any substance liberated at the surface of the muscle fiber in connection with the electric impulse propagation along the plasma membrane to activate the contractile elements by diffusion of the substance through the fiber, because such a mechanism would be too slow The activation is, furthermore, highly temperature-dependent, which also militates against a mechanism involving simple diffusion of an activating substance

There is good evidence that the transverse tubules (T tubules) of the sarcomere are involved in the propagation of impulses inward from the plasma membrane These tubules measure about 300 Å in diameter and run in a transverse direction through the muscle fiber They are called the T system to distinguish them from those tubules of the sarcoplasmic reticulum which are mainly longitudinal in orientation and, therefore, sometimes referred to as the L system The latter tubules form irregular networks along the myofibrils between the levels at which the tubules of the T system extend across the muscle fiber The tubules of the T system are arranged in a regular fashion extending from the plasma membrane toward the center of the muscle fiber at the Z line and/or at each A-I band boundry, there is substantial evidence that these tubules are in continuity with the plasma membrane and, therefore, that the narrow channel bounded by the tubule wall is continuous with the extracellular space It has already been noted that when a muscle fiber is stimulated locally by means of a microelectrode in contact with its surface, the threshold for a local response is minimal at the level of the I-A-band boundary in muscles where the T system is located at that level and at the Z line in muscles where the T system is confined to the Z line level In the latter case a localized contraction of the two half-sarcomeres on either side of the Z line can be elicited by applying a minimal stimulus

The nature of the impulse propagation along the T system is unknown Structurally, the conditions are fulfilled for an electrotonic spreading along the tubules of the T system of the electrical potential changes associated with the sarcolemmal membrane action potential

The activation of the contractile machinery ultimately depends on the effect of an activating factor capable of releasing the reactions that generate the active state It now appears that Ca^{++} ion is the final mediator of the process linking excitation of the plasma membrane to the initiation of contraction (excitation-contraction coupling), the action of Ca^{++} resulting from its binding to troponin (vide infra) The sarcoplasmic reticulum (S R) has been shown in vitro to be able to concentrate calcium The parts of the S R that are located closest to the A-I-band overlap are characterized by a high ATPase activity which may be related to a "calcium pump," the ATP splitting furnishing the energy to drive the pump In addition Ca may be reversibly bound to the S R It therefore appears likely that an impulse transmitted by the T system to the S R brings about a release of Ca that has been pumped into and or bound by the S R This release of Ca then initiates the activation of the contractile machinery

ELECTRICAL STIMULATION VIA THE SARCOLEMMA

Although the normal path of the activity is via the neuromuscular junction, it is possible to stimulate the sarcolemma directly If the intensity of the stimulus is subliminal a brief local depolarization can be induced which will reverse upon the removal of the stimulus However, if the stimulus depolarizes the membrane to the critical firing potential (threshold) an action potential will be observed Whereas the subliminal response is decremental and does not propagate far from the site of stimulation the action potential is self-regenerative and will be conducted in both directions away from the site of stimulation

Action potentials recorded from fast skeletal muscles depolarize the membrane rapidly, overshoot, and then repolarize rapidly to a level of approximately 70 mv (in the frog sartorius) At this level of membrane potential the repolarization process develops a relatively long time constant The membrane voltage during this latter period of slower repolarization is referred to as a negative after potential, as in the case of nerve The period of negative after potential in turn is followed by a period of hyperpolarization (the positive after potential) and then by a gradual return towards the normal resting potential, also similar to the response of nerve fibers The contribution of various sites on the sarcolemma (surface membrane, transverse tubules) to the ionic fluxes which determine the shape of the action potential has been investigated extensively, and it has been shown that depolarization of the muscle cell membrane is largely due to Na^{+} ion influx in most species although, in the crab, Ca^{++} ions are responsible for carrying most of the depolarizing charge into the cell

When the sarcolemma is depolarized or hyperpolarized by passing cathodal or anodal currents through it, it responds (like the neurilemma) as if it has both resistive and capacita-

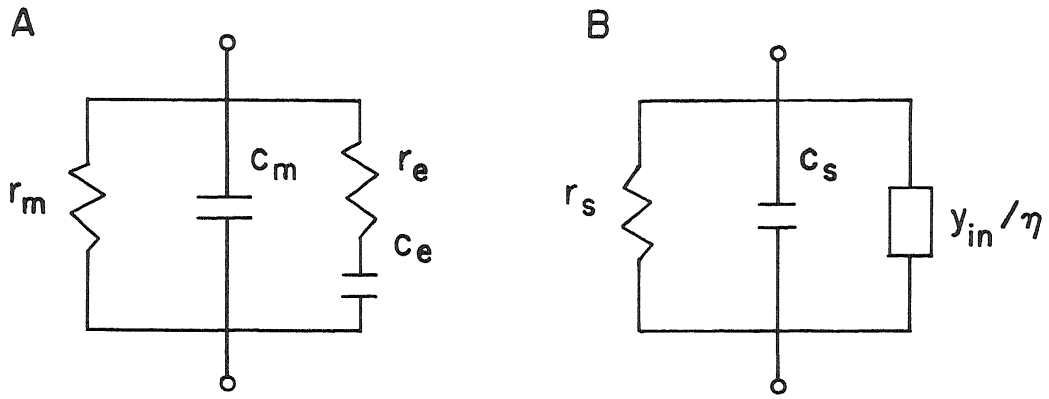


Fig 1 82 Two circuits considered for the element of inside-outside impedance in frog muscle fibers All parameters in both circuits refer to a unit length of fiber *A* four lumped parameter circuit r_m and c_m represent properties of the surface membrane and r_e and c_e properties of the path via the T system *B* distributed parameter circuit r_3 and c_3 represent properties of the surface membrane y_{in} is the input admittance of a lattice model for the T system and η is the sarcomere length (From Schneider 1970)

tive elements in parallel as well as in series Following a square wave stimulus, the membrane potential depolarizes exponentially with a time constant indicating the presence of capacitative elements which are accumulating charge The voltage drop across the membrane indicates resistance in the circuit The site and magnitude of these components have been investigated more thoroughly in nerve fibers, however, recent studies of frog sartorius have revealed additional components in muscle contributing to the electrical properties of the membrane in ways which have yet to be explained definitively

Various models have been proposed to explain the electrical properties of the muscle cell membrane For example, it has been suggested that the input impedance of frog muscle fibers may be compared to that of a cable having inside-outside impedance elements consisting of separate resistance and capacitance paths which are in parallel with a resistance and capacitance in series (Fig 1 82, model *A*) However, such models do not account for a radial potential gradient, inasmuch as the circuits are composed exclusively of lumped resistances and capacitances It has been suggested that radial gradients are possible only with circuits in which some of the resistance or capacitance elements are distributed throughout the fiber cross-section (Fig 1 82, model *B*) Thus, assuming that the contractile system is uniform across the fiber, only a system with spatially oriented resistance-capacitance elements can account for the observations that deep myofibrils are activated later than, and require greater surface membrane depolarization for activation than myofibrils located near the fiber surface

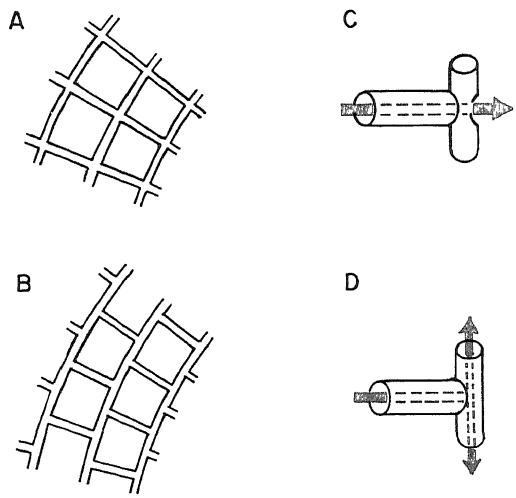


Fig 1 83 Geometries considered for the T system lattice *A* and *B* portion of the in register (*A*) or out of register (*B*) lattice *C* and *D* elements combined to give a single equivalent radial element Arrows indicate principal path of internal current for in register (*C*) or out of register (*D*) cases of the model (From Schneider 1970)

The possibility that there is a lattice system in the transverse tubules which might be represented by the model shown in Figure 1 83 was tested by stimulating muscle with AC current of varying frequency and then analyzing the (phase angle frequency) response of the fiber in terms of both of the models shown in Figures

1 82 and 1 83 It was found that the observed time constants and capacitance values were similar for both models (i e, C_e/C_m was 1 73 for model *B* and 1 74 to 1 8 for model *A*)

The calculated time constant r_m (C_e + C_m) was 17 8 msec using model *B* and 14 3 msec using model *A*

The fiber input impedance data exhibited a smoother transition in the intermediate frequency range when model *B* (which represents the T tubule lattice system) was compared to model *A* It is also noteworthy that the lattice structure accounts for the linear decrement of the applied current

Although the transverse tubules appear to be an extension of the sarcolemma, the membranes may have different conductance properties, as suggested by the analysis of the action potential recorded after the tubular system has been destroyed with hypertonic Ringer's-glycerol solution and the analysis of membrane currents after voltage clamping

The usual rapid repolarization that follows a spike generated by the muscle membrane is due to a delayed rectification of the membrane Inactivation of the conductance of Na^+ ions and increase in the conductance of K^+ ions are responsible for the quick phase of repolarization, following which there is an early and late negative after potential with the membrane remaining slightly depolarized for about 0 25 sec Then the membrane potential drops below the resting potential (hyperpolarization) and again returns to the resting level with a relatively long time constant

The origin of the fast phase of repolarization (delayed rectification) seems to be on the surface membrane, whereas, the after potential seems to originate mostly in the transverse tubules It has been shown that if the transverse tubules are destroyed by exposure to hypertonic solutions, muscle will no longer contract in response to stimulation (excitation-contraction uncoupling), in addition, the total membrane capacitance is reduced from 6 5 to 2 6 μF per cm ^{2} and the after potentials will disappear, although the earlier features of the action potential (spike and the rapid phase of repolarization) persist

The after potentials are believed to be due to K^+ ion accumulation in the tubules Using rubidium in place of potassium, it has been observed that the after potentials can be suppressed.

The contribution of Cl ions to the ionic currents of the muscle cell membrane have long been controversial It is now known that chloride conductance (G_{Cl}) is pH sensitive At values of pH of about 5 6 G_{Cl} is inhibited Under such conditions it has been shown that G_K in the tubules changes in direction during the after potential, this anomalous rectification seems to be a property of the surface membrane, whereas, the slow rectification (which accounts for the return from maximal hyperpolarization to the resting potential at the end of the action potential cycle) seems to take place within the tubules (which occupy a total of 0 002 to 0 003% of the fiber volume)

Many of the observations noted above have been verified by the recent observations of Adrian, Chandler, and Hodgkin (1970) to the effect that the potassium current can be divided into three separate components *1*) A current in a delayed rectifier channel of the muscle cell membrane which reaches a maximum in about 100 msec at a membrane potential of − 30 mv and then declines with a time constant of about 4 msec until the fiber is repolarized to − 100 mv This current inactivates completely *2*) A channel in which repolarization is associated with an increase in G_K, which is one or two orders of magnitude slower than the K current in the delayed rectifier channel *3*) A very slow inward K current which has a time constant of 0 25 sec at − 150 mv In addition a path for active uptake of K^+ ion is also postulated The data on current flow obtained with the voltage clamping technique also suggests that *1*) the delayed rectifier channel must be in the surface membrane, whereas, the two other channels must be in the T tubules, *2*) that G_K rather than tubular K concentration changes during the positive and negative after potentials, and *3*) that the after potentials can be explained if one supposes that potassium ions do not carry all the current or that some variable other than membrane potential is the factor which controls potassium permeability

RECAPITULATION OF THE ROLE OF MONO- AND DIVALENT IONS

1) K^+ ion, the major component of the intracellular fluid, functions chiefly to maintain the resting potential which approximates the K^+ equilibrium potential If the concentration of K in the extracellular fluid is increased, resting

membrane potential will decline according to the Nernst equation $E = RT/F \ln [K_I]/[K_E]$, where K_I is the intracellular K^+ concentration and K_E is the extracellular K^+ concentration

2) Na^+ ion, the major component of the extracellular fluid, functions to maintain the osmotic pressure (osmolality) of the extracellular compartment The permeability of the muscle cell membrane to Na^+ is very low at rest but increases markedly following stimulation Thus, Na^+ ions "rush" into the cell and these positive charges depolarize the membrane Because Na^+ permeability is voltage dependent, more Na^+ ions "rush" into the cell as depolarization progresses in what amounts to an autocatalytic, positive feedback process which is propagated along the muscle membrane This self-regenerative process is qualitatively identical to the action potential of nerve For a brief instant the inside is positive with respect to the outside At this point (the peak of the overshoot) the membrane potential approaches the Na^+ equilibrium potential However, the increased Na^+ permeability quickly reverses, permeability to K^+ increases and the repolarization process is initiated

The monovalent ion, Li^+, can replace Na^+ as a carrier of positive charge into the muscle fiber, however, Li^+ ions tend to accumulate inside the fiber because there is no active transport mechanism to remove Li^+ from the intracellular space

3) The role of Ca^{++} ions in muscle physiology will be discussed below in connection with the relaxation of muscle and under the heading of "Excitation-Contraction Coupling"

4) Mg^{++} ions are important for the ATPase activity of myosin Most of the Mg^{++} is bound to muscle components, although the precise distribution of Mg^{++} among these components has not yet been clarified

5) Cl^-, the major anion in the extracellular fluid, passively follows the cations moving across the muscle cell membrane distributing according to the Gibbs-Donnan equilibrium

Relaxing Factor

The relaxation of the muscle fiber is associated with a removal of Ca^{++} ions via a calcium pump and/or specific binding by the sarcoplasmic reticulum

When sarcotubular vesicles are isolated from muscle homogenates by centrifugation, they can cause isolated actomyosin systems to relax, an effect which is due to their ability to accumulate ionic calcium and thereby remove it from the medium

Addition of ATP to a hydrous gel of actomyosin shrinks the gel, reducing its water content This process (superprecipitation) is believed to be an elementary contraction process The same process can be seen in glycerinated muscle fibers (i e , muscle fibers extracted with 40 to 50% glycerin) When these fibers are exposed to ATP they show all the responses of a living contracting muscle However, when ATP is removed from the glycerinated fibers, they do not relax Clearly then, in these systems, a factor is missing which in living systems promotes relaxation following contraction

The ionic strength, ATP concentration, pH, temperature, and other conditions can be altered so that the superprecipitate of the actomyosin gel as well as the contracted glycerinated fibers can be made to relax However, in the living system all of the above conditions are probably constant, furthermore, the energy expenditure needed to effect changes in the concentration of salts, Mg^{++} or ATP of sufficient degree to promote relaxation precludes such mechanisms on thermodynamic grounds

The search for a physiological factor which reverses the contractile process has aroused considerable interest during the past 20 years In the early 50's, Marsh observed that a shrunken actomyosin gel could be made to take up water by the addition of a supernatant fluid which had been recovered from the centrifugation of ATP-treated muscle homogenate This muscle extract could relax contracted glycerinated muscle fibers The unknown component responsible for these effects was called "relaxing factor" Following the demonstration that relaxing factor was precipitated by 40% saturated ammonium sulfate, it was suggested that it was a soluble enzyme which interacted directly with myosin ATPase However, soon after it was found to sediment in a moderate gravitational field, and it therefore appeared to be particulate in nature Many investigators attempted to find a small molecule which, on release from the particulate, "relaxed" the ATP-contracted actomyosin system However, several years elapsed before it became apparent that the relaxing effect could be due to the removal of an element from the contracted actomyosin system, and therefore that the re-

laxing factor could be the system involved in the removal and sequestration of this element

In the mid-50's it was observed that the addition of ethylenediaminetetraacetic acid (EDTA) in the presence of ATP mimicked the physiological action of the relaxing factor, but it was not until 1959 that Weber demonstrated that the ATP-actomyosin system depended on minute amounts of Ca^{++} for its ATPase activity

The effect of the relaxing factor was then conceived to be a chelating system which was responsible for the removal of an agent, Ca^{++}, which regulated the actomyosin interaction The relaxing factor was identified on further study as the fragmented sarcoplasmic reticulum recovered from a muscle homogenate It was shown that these vesicular fragments could indeed concentrate Ca^{++} against very high gradients

The Regulatory Role of Calcium

It is known that pure actin and pure myosin can interact when ATP is present Ca^{++} ions, although not directly involved in this reaction, are needed for the release of an in vivo repression which prevents the interaction of actin and myosin

The site of storage of calcium is the sarcoplasmic reticulum (SR) The binding capacity of the SR system for Ca^{++} ions is 150 mM of Ca^{++}/g of protein with an association constant of $7 \times 10^6 M^{-1}$ The protein content of the SR is approximately 5 mg /ml of intact undiluted homogenized muscle The concentration of the calcium-receptive sites of the contractile system is about 0 1 mM, the association constant of calcium for the contractile system is 1×10^{-6} M^{-1} The concentration of rapidly-exchangeable calcium is 0 5 mM, and of nonexchangeable calcium 0 7 to 0 9 mM Most of the nonexchangeable calcium is bound to actin The concentration of free Ca^{++} ion in resting muscle is 3×10^{-7} M That the rate of binding of Ca^{++} ion to the SR system is indeed correlated with the rate of relaxation of muscle in vivo is shown by the finding that 40 mμM of calcium was bound by 1 mg of SR protein within 30 msec at 20° C This means that the SR content of 1 g of muscle can bind 0 2 moles Ca^{++} ions within 30 msec The amount of calcium released from 1 g of muscle ranges between 0 1 to 0 2 moles—in close accord, therefore, with the results of the binding studies However, because of the relatively high Ca^{++} ion concentration in the living cell it may take as long as 0 5 sec to remove 0 2 moles of calcium whereas the relaxation time of frog muscle at 20° C is only 0 1 sec This discrepancy may be explained by the suggestion of Ebashi and Endo (1968) to the effect that in the isolated system the SR vesicles may be partially inactivated and also that in vivo, there may exist a membrane potential across the SR tubules which enhances the removal of Ca^{++} ions from the sarcoplasm

SITE OF BINDING OF Ca IONS

Ca^{++} ions are bound to the SR membrane and concentrated in the lumen of the SR tubes It has been calculated that more than 90% of the calcium in the SR system is in bound form

In the absence of ATP, Mg^{++} binds readily with the SR system, however with addition of ATP, Ca^{++} replaces Mg^{++} It is suggested that the binding sites have a higher affinity for Mg^{++} than for Ca^{++} However, ATP may cause a conformational change in the relevant SR proteins, thereby increasing their affinity for Ca^{++}

Ebashi has suggested that Ca^{++} ions first bind with specific sites in the membrane of the SR system (step 1) and that this is the crucial step in the relaxation of the muscle Most of the Ca^{++} which acts as the trigger for contraction may be derived from this bound form Part of the initially bound Ca^{++} is then transported into the lumen (step 2) of the reticulum This transport is mainly regulated by the free Ca^{++} ion concentration inside the lumen which may not exceed 1 mM It is possible that a part of transported Ca^{++} is bound to certain structures in the lumen ATP is absolutely necessary for step 1 and its breakdown may take place in association with step 2

Whatever may prove to be the detailed mechanism of Ca^{++} accumulation it is clear that during the resting state the SR system is capable of reducing the sarcoplasmic Ca^{++} content below the threshold level for the activation of the contractile process (i e , less than 10^{-7} M)

Ca^{++} is the common regulatory ion in smooth and cardiac as well as skeletal muscle, however, contractile elements from various types of muscle are also sensitive in some degree to other alkaline earth metal ions For example, cardiac contractile elements have been found to have appreciable sensitivity to Sr^{++} and Ba^{++}, whereas skeletal and visceral smooth muscle are less sensitive to the latter

ions It is of interest therefore that cardiac troponin appears to be six times more sensitive to Sr^{++} than skeletal troponin

Ca^{++} AND THE INTENSITY OF THE ACTIVE STATE

There is a direct relationship between calcium ion concentration [Ca^{++}] and the intensity of the active state [I_{as}] inasmuch as the velocity of contraction (which parallels I_{as}) is proportional to [Ca^{++}]

Increasing free Mg^{++} ion concentration from 0 05 to 8 mM decreases the Ca^{++} sensitivity of the contractile system Although the total magnesium concentration in muscle amounts to 10 to 12 mM, the concentration of free Mg^{++}, is likely to be low, because myoproteins (200 mg /ml muscle homogenate) offer many binding sites to Mg^{++}

In glycerinated rabbit muscle, as compared with intact muscle, the sensitivity of the contractile elements to Ca^{++} is higher by 0 5 pCa unit, presumably because the native tropomyosin in these systems may be partially destroyed by glycerination (pCa refers to the negative logarithm of the Ca^{++} concentration)

THE SITE OF ACTION OF CALCIUM

If Ca^{++} ions are introduced by microiontophoresis at consecutive points along the surface of the sarcomere of glycerinated crab muscle [soaked in relaxing medium containing Mg-ATP and ethyleneglycoltetra-acetic acid (EGTA)], it is found that the site which requires the lowest iontophoretic current for the initiation of contraction is at the level of the A-I junction

As the micropipette containing the Ca^{++} solution is moved away from A-I junction, progressively stronger currents are needed to induce contraction Thus, it was concluded that Ca^{++} ions had to reach the zone of A-I overlap in order to contract the sarcomere in this species Essentially the same phenomenon has been demonstrated autoradiographically in frog muscle which was exposed to Ca^{45} and quickly frozen immediately after stimulation Using this technique it was demonstrated that the density of Ca^{45} is highest at the A-I junction directly after the initiation of contraction One to two seconds thereafter, most of the Ca^{45} was no longer observed at the A-I junction but rather was seen to be stored in the SR system or the extracellular compartment Thus, these experiments provide strong evidence indicating that the site of action of calcium is at the area of overlap of the thick and thin filaments

Excitation-Contraction Coupling

From the foregoing it is clear that the stimulation of the muscle, either through its nerve or directly, sets up a process which starts on the sarcolemma and ends in the contractile machinery

The pathway from the surface to the core of the fiber has been observed only recently It should be recalled that minute electrical currents at the level of the Z band in frog skeletal muscle induce local contractions and also that, in the frog, the transverse tubular system is located at the level of the Z lines—whereas, in the crab muscle, local stimulation is effective only at the A-I junction area and the transverse tubules are also found in this area It should also be recalled that when the T tubules are destroyed by exposure of muscle fibers to glycerin, electrical stimuli no longer elicit a contractile response Thus, it seems clear that the depolarization process initiated at the myoneural junction in the postjunctional membrane must then travel to the core of the myofibril through the T-tubule system

Electron micrographs have shown that the closed end of the T tubules are in close proximity (50 Å) to the terminal cisternae Thus, it is conceivable that the wave of depolarization traveling down the T tubule may trigger a new event involving the sarcoplasmic reticulum and leading to the contractile response

On the basis of extensive observations it has become clear during the last few years that the passage of the wave of electrical activity via the T tubules into the core of the myofibril induces a release of calcium ions which had been bound to the membranes of the sarcoplasmic reticulum or stored within it After release the calcium ions diffuse to and bind at calcium sensitive sites (troponin A) It has been suggested that the binding of calcium to troponin A allosterically removes an inhibiting effect of troponin B on the interaction of actin and myosin and, thus, calcium is the link which couples excitation (stimulation) to contraction (response)

The entire process can be abbreviated as follows Stimulus → Depolarization of the sarco-

lemma → Action potential initiated and propagated along the T tubules → calcium released from the SR system → Calcium ions diffuse and attach to the active site on troponin A → the inhibitory effect of troponin B on the interaction of actin and myosin is removed → actin filaments slide along myosin filaments shortening the sarcomere This series of events is fully reversible under normal conditions Following repolarization there is an increase in the uptake of calcium ions by the sarcoplasmic reticulum As this sequestration proceeds, more Ca^{++} ions dissociate from their binding sites on troponin A resulting in a restoration of actomyosin-inhibiting property of troponin B and, accordingly, in a return of the thick and thin filaments to the resting state (relaxation) Indeed Podolsky and Teichholz (1970) have evidence that the rate of muscle relaxation is proportional to the removal of calcium from the contractile protein

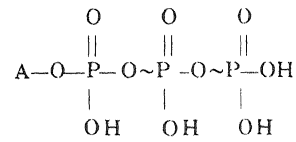
The Energy-Generating Systems

The energy for muscle contraction is derived from chemical reactions in the muscle fiber The main energy source is glycogen, which is present at a concentration of 0 5 to 1% of the wet weight of the tissue The energy associated with muscle contraction is made available without any oxygen being consumed, even when oxygen is present The oxidative (aerobic) chemical reactions are therefore not directly associated with muscle contraction but with the recovery processes which operate to provide energy in a form that is readily available to the contractile machinery, thereby securing a quick response to a stimulus.

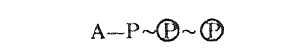
The compound that appears to be ultimately involved in providing energy to the contractile machinery is adenosine triphosphate (ATP) In this process, energy is utilized in a reaction coupled with the hydrolytic splitting of the terminal phosphate group from ATP,

$$ATP \rightarrow ADP + P_i \text{ (inorganic phosphate)}$$

This reaction involves a liberation of about 7600 to 7800 calories/mole It has become customary to refer to the bond between the terminal phosphate group and the neighboring phosphate group as a "high energy" or "energy rich" bond and to designate it by the symbol ~ The adenosine triphosphate structure, therefore, can be written as follows.



or



where A = adenosine adenine + D-ribose This formula shows two energy-rich phosphate bonds in the ATP molecule Only the one associated with the terminal group is directly utilized during muscle contraction

The ATPase activity of myosin permits the hydrolysis of ATP to take place at the contractile machinery itself where the conditions are topochemically favorable for energy transfer from ATP to the myosin-actin system

The restoration involves a phosphorylation of ADP to ATP In muscle, creatine phosphate (phosphocreatine) represents a second high-energy compound, in which phosphate is bonded to creatine by means of a high-energy bond Prompt resynthesis of ATP is achieved by phosphate group transfer from creatine phosphate to ADP, catalyzed by creatine kinase (Fig 1 84) Creatine P, depleted in this process, is regenerated by the transfer of a P from ATP to creatine During the recovery period ATP is formed by oxidative phosphorylation of ADP In this way a second form of storage of high-energy phosphate as creatine phosphate is built up in the muscle during recovery The ATP content of resting muscle is about 5×10^{-6} moles/gram muscle tissue which is four to six times less than the content of creatine phosphate The ATP and the creatine phosphate represent energy stored in a form that can be utilized rapidly Recharging of ATP through oxidative phosphorylation is a much slower process Since the high-energy phosphate storage in the muscle is sufficient to allow the muscle to contract only 50 to 100 times, the energy reservoirs are rapidly depleted For maintenance of muscle contractility, a steady supply of energy is required to recharge ADP

Since the reactions associated with the contraction, dephosphorylation of ATP, as well as the phosphate transfer from creatine phosphate to ADP, are nonoxidative processes, it becomes obvious that no oxygen is consumed during contraction provided that no recovery metabolism is maintained in parallel Oxygen is con-

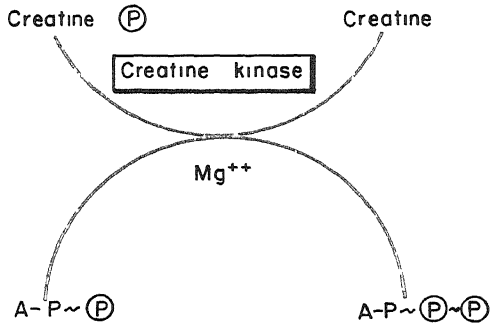


Fig 1 84 Resynthesis of ATP with phosphate from phosphocreatine

sumed during the recovery phase in connection with oxidative phosphorylation of ADP

The energy required for the high-energy phosphate bonds of ATP and creatine phosphate is supplied by glycolysis and by the oxidation of acetyl CoA (derived from pyruvic acid) in the citric acid cycle with the associated oxidative phosphorylation by the respiratory chain

In addition to glycogen, fatty acids and amino acids can furnish fuel for the muscle, and in resting muscle, glycogen breakdown is not required to maintain metabolism Muscle metabolism can be maintained by the uptake of carbohydrates and other compounds from the blood In resting muscle, carbohydrates are responsible for about 60% of the energy requirements of the tissue

When 1 mole of glucose is phosphorylated to glucose-6-phosphate and then enters the glycolytic pathway to be converted to lactic acid there is a net gain of 2 moles of ATP However, when glycogen per se is broken down to lactic acid, there is a net gain of 3 moles of ATP per glucose residue of glycogen so converted It is obvious that a limited recharging of ADP and creatine can occur in muscle without access to oxygen Breakdown of lactic acid, however, depends on the presence of oxygen, and under anaerobic conditions, the lactic acid concentration will increase in the muscle and then be transported by the blood to the liver

Glycolysis can be inhibited by iodoacetate poisoning and under those conditions the muscle will contract until the ATP and the creatine phosphate stores have been depleted

Under aerobic conditions, the end product of glycolysis is pyruvic acid instead of lactic acid Pyruvic acid is oxidatively decarboxylated to a derivative of acetic acid, acetyl coenzyme A (acetyl CoA), and enters the aerobic citric acid cycle of Krebs, which results in its complete oxidation to CO_2 and H_2O During the degradation, 15 ADP molecules are phosphorylated per molecule of pyruvic acid, which means that 30 ADP molecules are phosphorylated during oxidation of the two pyruvic acid molecules derived from each glucose molecule Since, as noted above, 3 ADP molecules are phosphorylated during glycolytic conversion to lactic acid of each glucose residue derived from glycogen and 6 ATP molecules derived from the oxidation of the 2 DPNH molecules generated during glycolysis, a total of 39 moles of ATP are formed from glycogen per mole of glucose residue degraded

Total oxidation of glucose to CO_2 and H_2O yields 686,000 cal /mole, but only 56,000 cal /mole is produced when the glucose degradation does not proceed further than to lactic acid The efficiency of glycolysis, that is, of the energy-generating reactions under anaerobic conditions, can be calculated by assuming that the phosphorylation of three ADP molecules represents about 24,000 cal This is $24,000/56,000 \times 100$, or about 43% of the potentially available energy during glycolysis following glycogenolysis In comparison with the energy released during complete oxidation of glucose—686,000 cal /mole—the energy stored in 3 molecules of ATP will represent $24,000/686,000 \times 100$, or about 3% of the total energy available as chemical energy in glucose This shows the low energy yields of anaerobic energy metabolism. In comparison, 36 + 3 ADP molecules are phosphorylated in the aerobic degradation of a glucose equivalent derived from glycogenolysis representing a free energy storage of about 312,000 calories The efficiency of the oxidative degradation of glucose is therefore $312,000/686,000 \times 100$, or about 45%, which is in good agreement with the efficiency estimated from measurements of the recovery heat

GLYCOGEN METABOLISM

The first step in the synthesis of glycogen in muscle involves phosphorylation of glucose to glucose 6-phosphate, a reaction catalyzed by *glucokinase* and associated with utilization of energy-rich phosphate in the form of ATP

Glucose + ATP → glucose 6-phosphate + ADP

The product, glucose 6-phosphate, is a relatively energy-poor compound and this reaction therefore means that the equilibrium state of this reaction favors the formation of glucose-6-phosphate (G-6-P). G-6-P is converted to glucose 1-phosphate, a reaction catalyzed by the enzyme *phosphoglucomutase* (Fig. 1-85). Glucose 1-phosphate is then activated by reacting with uridine triphosphate (UTP) to form uridine diphosphate glucose (UDPG). This reaction is catalyzed by *UDPG pyrophosphorylase*. The glucose moiety of UDPG is then transferred to a pre-existing glycogen molecule (primer) to which it is attached in an α 1, 4-linkage. This reaction is catalyzed by the enzyme *glycogen synthetase* (UDPG-glycogen-transglucosylase). The formation of such bonds gives rise to linear changes of glucose residues. At intervals chains of 8 to 12 glucose residues are picked up and transferred to make a branch which is attached to the main chain of glucose residues by means of an α 1,6 linkage. This reaction is catalyzed by the "*branching enzyme*" (amylo-1,4-1,6 transglucosylase). In this way, a branch point is established, and the glycogen molecule grows to a large, highly branched structure (Fig. 1-86) with a molecular weight that ranges from 240,000 to 10^7.

During glycogenolysis (Fig. 1-87), glycogen is first partially degraded to α-D-glucose-1-phosphate, this degradation is catalyzed by *glycogen phosphorylase*, an enzyme whose activity is confined to the 1,4 links. All of the residues on the branches, except for the proximal residue, are transferred to the main chain by the *debranching enzyme*. The 1,6 linkage involving the proximal residue is then split by *amylo-(1,6)-glucosidase*, releasing this branch point residue as free glucose. Thereafter the remaining molecule can be degraded completely by glycogen phosphorylase.

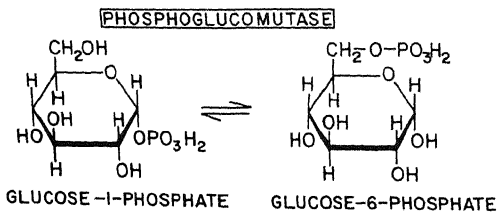


Fig. 1-85 Interconversion of glucose phosphates by the enzyme phosphoglucomutase.

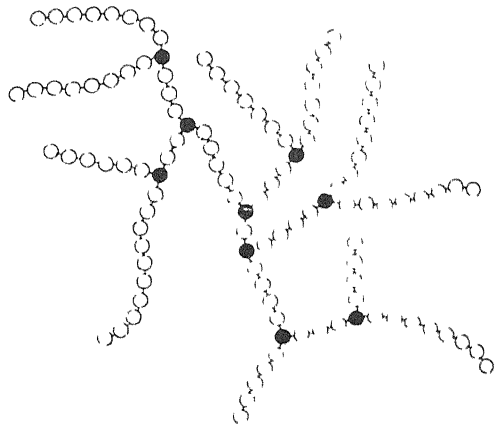


Fig. 1-86 Part of a branched glycogen molecule with each circle representing one glucose residue.

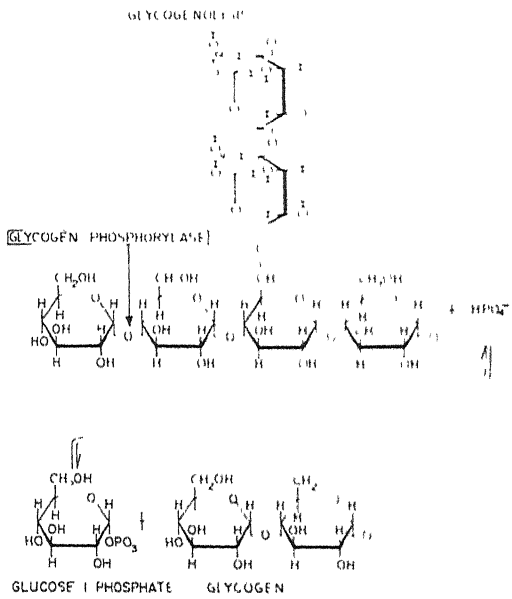


Fig. 1-87 Glycogenolysis.

Glycogen phosphorylase, with a molecular weight of 370,000, consists of four polypeptide chains. The molecule is acted on by a muscle enzyme, *phosphorylase phosphatase* to remove phosphate groups from serine residues present in each polypeptide chain. This results in a rearrangement of these chains to a dimeric form of the enzyme which is virtually inactive in the absence of adenosine-5'-monophosphate (AMP). The tetramer, which does not require AMP, is referred to as phosphorylase *a* and the less active dimer as phosphorylase *b*. The enzyme contains pyridoxal phosphate bound to a

lysyl residue through an azomethine linkage (Schiff base) Phosphorylase *b* is found in resting muscle Active enzyme, i e , phosphorylase *a*, is generated by the action of phosphorylase *b* kinase in the reaction

2 phosphorylase *b* + 4 ATP → phosphorylase *a* + 4 ADP

Phosphorylase *b* kinase (PBK) also appears in an active (phosphorylated) and an *inactive*[1] (*dephosphorylated*) form, the inactive form is converted to the active enzyme by an ATP coupled enzyme system, known as phosphorylase *b* kinase kinase (PBKK) PBKK, like many other protein kinases, requires adenosine 3′-5′-phosphate (cyclic AMP) for its activation

The breakdown of muscle glycogen is largely determined by the extent of conversion of phosphorylase *b* to phosphorylase *a* in a process which is under control of the nervous and endocrine systems via neurohumoral-hormonal entities such as epinephrine and by intracellular mediators ("second messengers") such as cyclic AMP Recently it has been found that the kinase kinase also catalyzes the conversion of glycogen synthetase to the inactive phosphorylated synthetase (glycogen synthetase D) and may be related to or identical with glycogen synthetase kinase, therefore, when the phosphorylase system is activated by phosphorylation, the synthetase is inactivated by the very same process This biochemical switch permits glycogen breakdown to occur without the concurrent competition of synthesis Such a coupling makes eminent physiological sense because futile competition between glycogen synthesis and breakdown with a consequent wasteful dissipation of energy is thus prevented

Glucose 1-phosphate is converted to glucose 6-phosphate in a reaction catalyzed by the enzyme *phosphoglucomutase*

Glucose 1-phosphate⇌glucose 6-phosphate

This reaction requires glucose-1,6-diphosphate and Mg^{++} as cofactors The enzyme is reversibly phosphorylated during this reaction by a transfer of one of the phosphate groups from glucose-1,6-diphosphate to a serine residue The phosphate group of the esterified enzyme then presumably is transferred to the 6-carbon in glucose 1-phosphate to regenerate glucose 1,6-diphosphate The glucose-1,6-diphosphate necessary for this reaction is formed by transfer of a phosphate group from ATP to glucose 1-phosphate in a reaction catalyzed by *phosphoglucokinase* These reactions can be summarized as follows

(1) glucose-1,6-diP + enzyme ⇌ glucose-6-P + enzyme-P

(2) enzyme-P + glucose-1-P ⇌ glucose-1,6-diP + enzyme

The resultant of these reactions is the conversion of glucose-1-phosphate to glucose-6-phosphate

Since glucose 1,6-diphosphate is regenerated in the above reactions the energy expenditure associated with them is small in relation to the net energy transferred to ATP during oxidation of glucose

The glucose 6-phosphate enters the main pathway of glycolysis followed by complete oxidation via the citric acid cycle and its associated oxidative phosphorylation system The scheme in Figure 1 88 summarizes the pathways of glycogenesis and glycogenolysis

GLYCOLYSIS

Glycolysis involves the degradation of glucose to pyruvic or lactic acid This breakdown of glucose can occur without access to oxygen, in which case the end product is lactic acid Two molecules of ADP are phosphorylated to ATP during glycolysis

Glucose + 2ADP + 2 P⇌2 lactic acid + 2 ATP

The conversion of free glucose to glucose-6-phosphate is bypassed when glycolysis follows glycogenolysis because glucose-6-phosphate is then formed directly from glucose 1-phosphate in a reaction not involving utilization of phos-

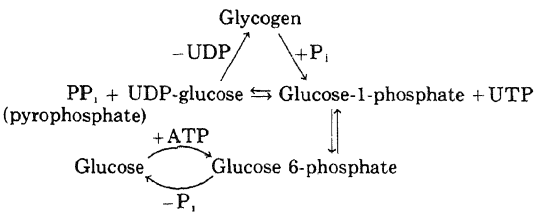


Fig 1 88 Summary of pathways of glycogenesis and glycogenolysis

[1] The dephospho form of PBK can be activated by Ca^{++} ions, a phenomenon which suggests that the enzyme may be rendered active in conjunction with the Ca^{++}-dependent triggering of the contractile process

phate-bond energy The net gain of glycolysis will therefore, in this case, be 3 moles of ATP per mole of glucose residue in glycogen

Glucose 6-phosphate is converted to fructose 6-phosphate in a reaction catalyzed by *phosphohexose isomerase* A second phosphate group is introduced at the 1-carbon of fructose 6-phosphate to yield fructose 1,6-diphosphate The phosphate group is transferred from ATP and the transfer is catalyzed by the enzyme *phosphofructokinase* Since 1 mole of ATP is dephosphorylated per mole fructose 1,6-diphosphate formed, this reaction involves expenditure of energy

In the next step the 6-carbon compound fructose 1,6-diphosphate is split between C-3 and C-4 to yield two three-carbon compounds, dihydroxyacetone phosphate and D-glyceraldehyde 3-phosphate The enzyme *aldolase* catalyzes this cleavage

The two triose-phosphates formed are interconvertible by isomerization, and dihydroxyacetone phosphate is converted to D-glyceraldehyde 3-phosphate in a reaction catalyzed by *triose phosphate isomerase*

D-glyceraldehyde 3-phosphate is then oxidized to 1,3-diphosphoglyceric acid by *phosphoglyceraldehyde dehydrogenase* (triose phosphate dehydrogenase) in the presence of inorganic phosphate, and DPN^+ bound to each subunit of the enzyme Sulfhydryl groups in the molecule are of decisive importance for the activity of this enzyme The inhibition of glycolysis by iodoacetate is attributed to the fact that this compound reacts with these sulfhydryl groups

The energy of oxidation involved in the oxidation of 3-phosphoglyceraldehyde to a carboxylic acid is retained within the molecule in the form of the bond energy of the intermediate, 1,3-diphosphoglyceric acid

The bond energy is sufficient for a transfer of a phosphate group from the 1 position of 1,3-diphosphoglyceric acid to ADP The enzyme *3-phosphoglyceric acid kinase* catalyzes this reaction

$$\text{1,3-diphosphoglyceric acid} + \text{ADP} \rightleftharpoons \text{3-phosphoglyceric acid} + \text{ATP}$$

The sum of this reaction and that involving oxidation of glyceraldehyde 3-phosphate is the following

$$\begin{array}{c}\text{Glyceraldehyde 3-phosphate} + P_i + DPN^+ + \text{ADP} \\ \Downarrow \\ \text{3-phosphoglyceric acid} + \text{DPNH} + \text{ATP} + H^+\end{array}$$

One molecule of ATP is thus formed for each molecule of 1,3-diphosphoglyceric acid dephosphorylated The energy now stored in the ATP molecule has been derived from the energy of the exergonic oxidation of D-glyceraldehyde 3-phosphate This is an example of the coupling of an energy-releasing oxidation with phosphorylation of ADP in order to retain energy in a chemically available form thus preventing it from being dissipated as heat Oxidation of diphosphopyridine nucleotide (DPNH) is also coupled to ADP phosphorylation and yields 3 moles of ATP per mole DPNH (see below)

An exchange reaction between cytoplasmic and intramitochondrial DPNH is required because the permeability of the mitochondrial membrane for DPNH is low Several mechanisms for such an exchange have been proposed, but the actual reactions that predominate in vivo are not known

The 3-phosphoglyceric acid is transformed into 2-phosphoglyceric acid in an intermediate step in which 2,3-diphosphoglyceric acid is utilized as a cofactor The overall reaction is catalyzed by *phosphoglyceromutase* and is similar in mechanism to that involving G-1,6-diP (see above)

2-Phosphoglyceric acid is converted by dehydration to phosphoenol-pyruvic acid in the presence of *enolase* The phosphate group of phosphoenol pyruvic acid is linked to the molecule by a high-energy bond. A transfer to ADP of the phosphate group yields 1 molecule of ATP and pyruvic acid This reaction is catalyzed by *pyruvate kinase* in the presence of Mg^{++}

This is the last reaction during glycolysis that can furnish energy for synthesis of ATP Under anaerobic conditions, DPNH formed as a result of conversion of 3 phosphoglyceraldehyde to 3-phosphoglyceric acid can be oxidized to DPN^+ in a reaction where pyruvic acid is reduced to lactic acid The coupling of these two reactions is essential under anaerobic conditions, because DPNH cannot be oxidized by the mitochondrial electron transport system which requires oxygen The coupling occurs, however, also in the contracting muscle with normal blood supply when the requirements regarding the speed of DPNH oxidation are too high to be satisfied by the mitochondrial system. Lactic acid is therefore formed also under aerobic conditions when the muscle activity exceeds a certain intensity

The coupling of the reactions leading to res-

toration of DPN$^+$ can be summarized as shown in Figure 1 89 No oxygen, however, is consumed during glycolysis and the presence of oxygen is not required

The reactions in the glycolytic breakdown of glucose are summarized in Figure 1 90

Glycolysis occurs in the sarcoplasm, and the enzymes involved are part of the easily extracted "soluble proteins" from muscle tissue which constitute about 30% of the total muscle protein

The *citric acid cycle* (tricarboxylic cycle or Krebs cycle) represents a cyclic sequence of chemical reactions, by which acetic acid bound to coenzyme A (CoA) as acetyl coenzyme A (Acetyl CoA) is oxidized to CO_2 and H_2O (Figs 1 91 and 1 92) Pyruvic acid derived from glycolysis furnishes one source of acetyl CoA Pyruvic acid is oxidized in a series of reactions involving several enzymes, and four cofactors which form a complex constituting *pyruvic acid dehydrogenase* The following overall scheme illustrates the net effect of the reactions of Figure 1 93

The acetyl CoA is an energy-rich compound and can also be formed from fatty acids and certain amino acids The citric-acid cycle, therefore represents a common pathway in the oxidative degradation of carbohydrates, fats, and proteins Glucose and fatty-acid metabolism represent, however, the main sources of acetyl CoA

In the citric-acid cycle, acetyl CoA reacts with the 4-carbon oxaloacetic acid to form the 6-carbon citric acid, a reaction catalyzed by citrate synthase, formerly referred to as the "*condensing enzyme*" Citric acid is converted to its isomer, isocitric acid Oxidative decarboxylation of isocitric acid results in the formation of α-ketoglutarate In a series of reactions analogous to the pyruvate dehydrogenase reactions (see above), succinyl CoA is formed from α-ketoglutarate and CoA The energy of the thioester bond of succinyl CoA is used to form GTP from GDP in a reaction producing free succinate and free CoA Succinate is oxidized to form fumarate and reduced FAD Hydration of fumarate to malate and finally oxidation of malate to form oxaloacetate completes the cycle One molecule of the 2 carbon acetyl fragment is oxidized during the cycle, with 2 molecules of CO_2 and 4 hydrogen atoms

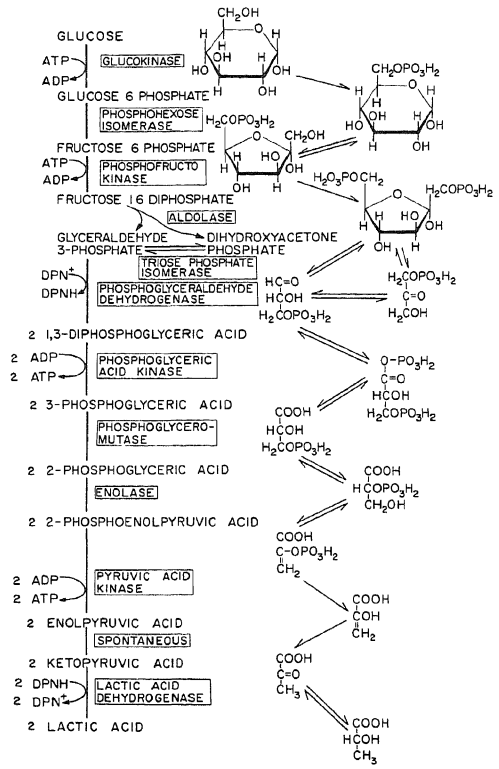


Fig 1 90 Glycolysis

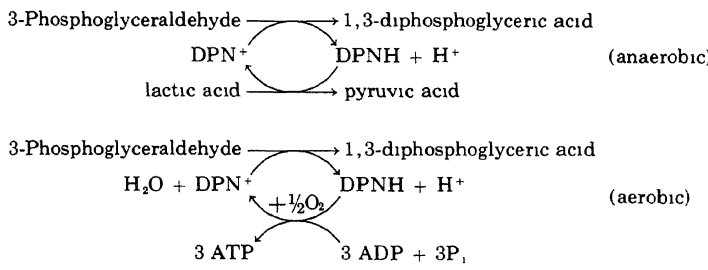


Fig 1 89 Anaerobic and aerobic restoration of DPN$^+$

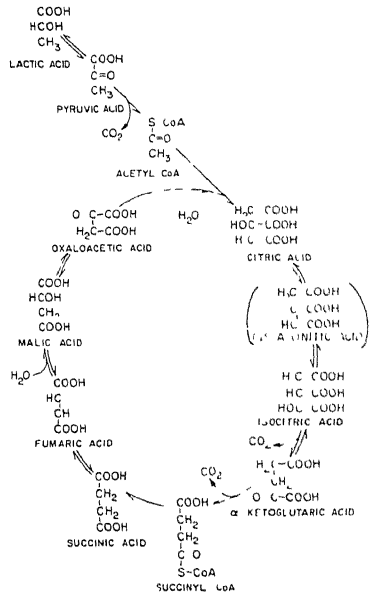


Fig 1 91 The citric acid cycle

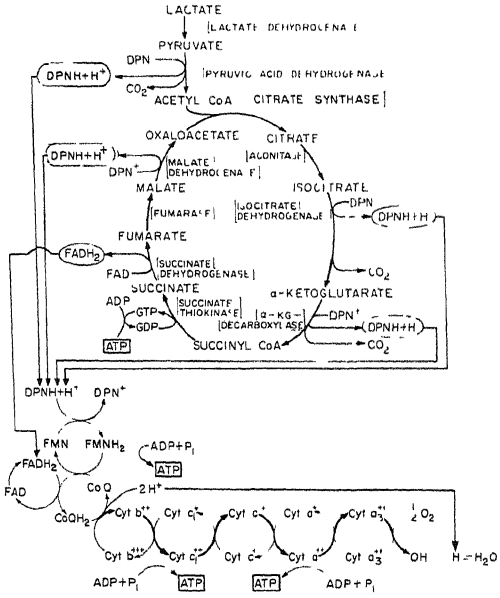


Fig 1 92 The citric acid cycle and the electron transport chain

removed The net exchange during one cycle is shown in the following scheme

$$CH_3COOH + 2O_2 \rightarrow 2CO_2 + 2H_2O + \text{energy}$$

The five energy yielding reactions involve 3 steps in which DPNH is produced, one step in which FADH is produced and one substrate level phosphorylation as follows

(1) Isocitrate + DPN → α ketoglutarate + CO_2 + DPNH + H
(2) α ketoglutarate + DPN + CoA → succinyl CoA + CO_2 + DPNH + H
(3) Succinyl CoA + GDP + P_i → succinate + CoA + GTP
(4) Succinate + FAD → fumarate + FADH
(5) Malate + DPN → oxaloacetate + DPNH + H

OXIDATIVE PHOSPHORYLATION (FIG 1 92)

The reactions in the citric acid cycle are coupled to the respiratory chain consisting of a sequence of hydrogen and electron carriers

The electron carriers are iron containing hemoproteins or cytochromes which transfer electrons by alternating between reduced and oxidized states, the iron switching between ferrous and ferric forms An intermediate step involving the removal of hydrogen from DPNH by DPNH dehydrogenase (which is a flavoprotein), precedes the passage of electrons along the series of cytochromes in the carrier chain The flavoprotein is oxidized by ubiquinone (coenzyme Q, CoQ), a lipid-soluble quinone acting as a coenzyme The exact position of ubiquinone in the respiratory chain is not known, but it is frequently considered to be positioned between FAD and cytochrome *b*, where it then would act as a hydrogen acceptor In the next step, involving oxidation of ubiquinone, the electron from hydrogen is transferred from ubiquinone to the first cytochrome in the cytochrome series, cytochrome *b* (Cyt *b*), and the proton is added to the medium This, therefore, is the step at which hydrogen transfer changes to electron transfer in the respiratory chain Electrons are now passed along the chain of cytochromes, involving in addition to cytochrome *b*, cytochrome c_1 (Cyt c_1), cytochrome c (Cyt c), cytochrome *a* (Cyt *a*) and cytochrome oxidase or cytochrome a_3 (Cyt a_3) Only the last member or the last two members of the group, cytochrome a_3 and *a*, can react directly with oxygen When oxygen is reduced by cytochrome oxidase, hydroxyl ions are formed that combine with the protons generated at the oxidation of ubiquinone to form water The net effects of the electron transport along the respiratory chain are illustrated in Figure 1 92

During the electron transfer along the respiratory chain, starting with DPNH, 3 ADP mol-

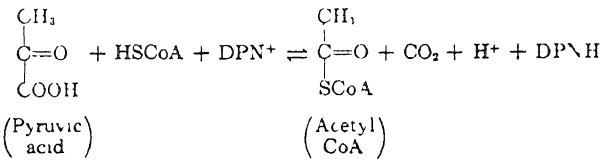


Fig 1 93 Summary of reactions in which acetyl CoA is produced as pyruvic acid is oxidized

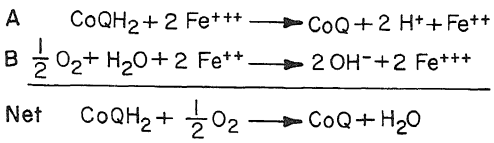


Fig 1 94 Summary of reactions in which ubiquinone ($CoQH_2$) oxygen and cytochrome oxidase interact to form water

ecules are phosphorylated to yield 3 ATP molecules per atom of oxygen This is expressed as a P/O ratio of oxidative phosphorylation of 3 The energy for the formation of the high-energy bonds is derived at three points along the respiratory chain, these three points are considered to be in the segment between DPNH and cytochrome b, between cytochromes b and c, and at the oxidation of cytochrome *c* by cytochrome oxidase The minimum difference in oxidation-reduction potential which is required to provide about 8000 cal for phosphorylation of ADP is 0 18 v A difference in potential that is sufficiently large to satisfy this requirement exists at the first point of phosphorylation, and this reaction can therefore proceed in connection with the transfer of a single electron The potential differences at the two other points are, however, too small to make energy available for ATP synthesis in connection with the passage of a single electron Two electrons would be required It is not understood how the energy associated with two consecutive single electron passages can converge to generate 1 molecule of ATP

At the fourth energy-yielding reaction in the citric-acid cycle described above, succinate is oxidized to fumarate by succinate dehydrogenase Since the oxidation of DPNH is bypassed, the P/O ratio is 2 instead of 3

The mechanism by which the phosphorylation of ADP is coupled to the electron transport along the respiratory chain is unknown It has been assumed that the energy is transferred to ATP through two intermediate compounds via energy-rich bonds If this coupling mechanism is inhibited by drugs, respiration can, under experimental conditions, proceed without the generation of ATP, indicating that the electron transport system can remain intact when the phosphorylation is uncoupled

Under physiological conditions, however, respiration and phosphorylation are tightly coupled and the rate-limiting factors in these processes are the concentrations of ADP and inorganic phosphate The ADP concentration, on the other hand, is determined by the rate at which ATP is utilized This represents a self-regulatory mechanism by which the rate of energy generation is adjusted to the requirements of the cell

Table 1 6 summarizes the energy-yielding and energy-requiring reactions during glucose degradation In reactions 4, 5, and 9, ATP is formed from ADP participating in the reaction In the other reactions except for reaction 10, ATP generation is a consequence of DPNH oxidation in the respiratory system of mitochondria with phosphorylation of ADP coupled to the reaction

$$DPNH + H^+ + {}^1\!/_2O_2 = DPN^+ + H_2O$$

DPN is bypassed in reaction 10 and hydrogen is transferred to FAD, and ATP is formed in connection with a coupling of $FADH_2$ oxidation and ADP phosphorylation according to the following net reaction

$$FADH_2 + {}^1\!/_2O_2 \rightleftharpoons FAD + H_2O$$

In reactions 6, 7, and 8 1 mole of CO_2 is formed for each mole of DPNH formed In glycolytic degradation of glucose equivalents derived from glycogenolysis reaction 1 is not involved and the total number of moles of ATP formed per mole glucose equivalents is, therefore, 39

The enzymes of the citric-acid cycle and the components of the respiratory chain are located in the mitochondria of the muscle cells While the former enzymes can be extracted fairly easily from mitochondria, the components of the respiratory chain are tightly bound to the mitochondrial structure

Table 1 6
Energy yield in glucose degradation

Reaction	ATP yield
(1) Glucose + ATP → glucose 6-phosphate + ADP	1
(2) Fructose 6-phosphate + ATP → fructose 1 6-diphosphate + ADP	1
(3) 2 D-glyceraldehyde 3-phosphate + 2 DPN$^+$ + 2 P_i ⇌ 2 1, 3 diphosphoglyceric acid 2 DPNH + 2 H$^+$	6
(4) 2 1,3-diphosphoglyceric acid + 2 ADP ⇌ 2 3-phosphoglyceric acid + 2 ATP	2
(5) 2 Phosphoenolpyruvic acid + 2 ADP → 2 pyruvic acid + 2 ATP	2
(6) 2 Pyruvic acid + 2 CoA + 2 DPN$^+$ → 2 acetyl CoA + 2 CO_2 + 2 DPNH + 2 H	6
(7) 2 Isocitric acid + 2 DPN$^+$ ⇌ 2 α ketoglutaric acid + 2 CO_2 + 2 DPNH + 2 H$^+$	6
(8) 2 α-ketoglutaric acid + 2 CoA + DPN$^+$ → 2 succinyl CoA + 2DPNH + 2 H$^+$ + 2 CO_2	6
(9) 2 Succinyl CoA + 2 GDP + 2 P_i → 2 succinic acid + 2 CoA + 2 GTP 2 GTP + 2 ADP ⇌ 2 GDP + 2 ATP	2
(10) 2 Succinic acid + 2 FAD ⇌ 2 fumaric acid + 2 $FADH_2$	4
(11) 2 Malic acid + 2 DPN$^+$ ⇌ 2 oxalacetic acid + 2 DPNH + 2 H	6
Total	38

As noted in Chapter 1, the mitochondrion is characterized by a double surface membrane which forms a boundary separating the interior of the mitochondrion from the surrounding cytoplasm In the interior, a system of inner membranes forms plates that are arranged more or less in parallel, and extend perpendicular to the long axis of the mitochondrion, almost across its whole diameter Sometimes these inner membranes are arranged diagonally or longitudinally In the muscle mitochondria, the inner membranes are closely arranged and separated by layers of mitochondria matrix that are only a few hundred Ångstrom units in thickness (Fig 1 95)

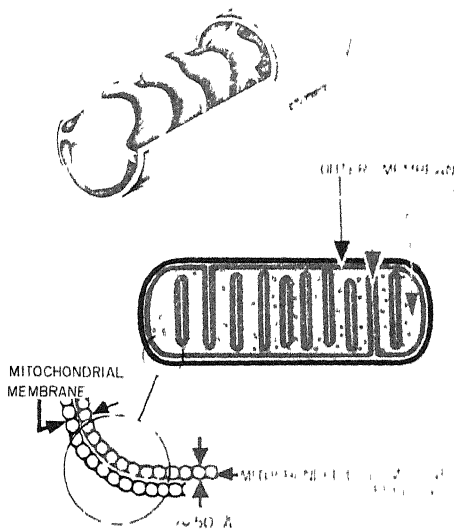


Fig 1 95 The structure of mitochondria

The mitochondrial membranes consist of two presumably identical elements that are in close apposition to form a complex membrane structure The two membrane elements can easily be separated by uptake of water and they then appear to bound vesicular structures The whole mitochondrion also can increase its volume considerably by uptake of water, and the maintenance of a normal volume depends on retaining the metabolism of the mitochondrion The mitochondria contain 30 to 40% of lipid by weight, mainly located in the membrane structures

It is generally accepted that the components of the respiratory chain are arranged in the proper sequence in the membrane elements of the mitochondria and prevented from random motion due to this association Another type of constraint which may be imposed upon the mitochondrial membrane is the orientation of the binding sites of the carriers; e.g , the binding sites for DPHN dehydrogenase and succinic acid dehydrogenase are assumed to face the matrix at the surface of the membrane elements It has been speculated that the electron transfer might be facilitated by the close arrangement of the components of the respiratory chain, which would allow transfer with small changes in orientation of the individual electron carriers

Electromyography

The structural basis of electromyography is the motor unit In normal skeletal muscles,

fibers probably never contract as isolated individuals Instead, several of them contract at almost the same moment, all being supplied by branches of the axon of one spinal motor neuron

FUNCTION OF A MOTOR UNIT

The muscle fibers of a motor unit normally contract sharply upon the arrival of impulses at frequencies ranging up to 50/sec This frequency seems to be the upper physiological limit for axonal propagation of motor neurons in most mammals including man However, motor units may be deliberately fired at much slower frequencies, even to as low as single isolated contractions at will Normal human beings can learn easily to provide various frequencies of impulses usually below 16/sec — from their spinal motor neurons

The number of striated muscle fibers that are served by one axon, i e , the number in a motor unit, varies widely Generally, muscles controlling fine movements and adjustments (such as those attached to the ossicles of the ear and of the eyeball, larynx, and pharynx) have the smallest number of muscle fibers per motor unit The muscles that move the eye have less than 10 fibers per unit, the muscles of the middle ear, 10 to 125, the laryngeal muscles 2 to 3, and the pharyngeal muscles have 2 to 6 These are all rather small and delicate muscles and they control fine or delicate movements On the other hand, large, coarse-acting muscles have motor units with many muscle fibers, e g , the human gastrocnemius has 2000 or more

Even the largest bundles of muscle fibers are quite small, and so a strong contraction of a skeletal muscle requires the participation of many motor units Further, there is a complete asynchrony of the motor unit contractions imposed by asynchronous volleys of impulses coming down the many axons Thus, with motor units contracting and relaxing at differing rates, a smooth pull results (In certain disturbances the contractions become synchronized, resulting in a visible tremor)

The fibers in a motor unit of the rabbit's sartorius may be scattered and intermingled with fibers of other units Thus, the individual muscle bundles seen in routine histology do not correspond to individual motor units as such In rat muscle, the fibers of a motor unit are widely scattered In man, a similar condition is probable, the spike potentials of each motor unit in the biceps brachii are localized to an approximately circular region of 5-mm diameter in which the fibers of the unit are confined However, the potentials can be traced in their spread to over 20-mm distance, thus the area of 5-mm diameter includes many overlapping motor units

Motor Unit Potential. During the normal twitch of a muscle fiber, a minute electrical potential is generated, which is dissipated into the surrounding tissue The duration of the action potential associated with this twitch is about 1 to 2 msec , or even 4 msec Since all the muscle fibers of a motor unit do not contract at exactly the same time—some being delayed for several milliseconds—the electrical potential developed by the single twitch of all the fibers in the motor unit is prolonged The electrical result of the motor-unit twitch, then, is an electrical discharge lasting about 5 to 8 msec (and often as long as 12 msec) The majority of these motor-unit potentials have an amplitude of around 0 5 mv When displayed on a cathode-ray oscilloscope, the result is a sharp spike that is most often biphasic (Fig 1 96), but it may also have a more complex form, depending on physical and other factors

Generally, the larger the motor-unit potential registered, the larger is the motor unit producing it However, complicating factors, such as distance of the unit from the electrodes and the types of electrodes and equipment used, determine the final size and pattern of the individual motor-unit potential that is recorded For further details, the reader should consult

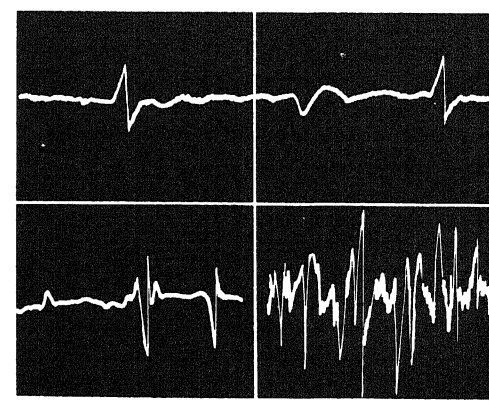


Fig 1 96 Normal electromyograms The single potential in the upper left corner has a measured amplitude of 0 8 mv and a duration of 7 msec

the papers of Basmajian (1967), Buchthal et al (1957), and Hakansson (1957)

Petersen and Kugelberg (1949) have shown that the electrode type affects the recorded duration and amplitude of the action potentials They demonstrated characteristic variations, e g , the smallness of potentials in facial muscles as compared with those in muscles of the extremity Furthermore, they (and others) have found a slight prolongation with advancing age

Under normal conditions, the smaller potentials appear first with a slight contraction As the force is increased, larger and larger potentials are recruited, this being the normal pattern of recruitment This normal pattern is absent in cases of partial paralysis due to injuries or lesions of the lower motor neuron, i e , the small potentials never appear, apparently because only the larger motor units have survived

Electromyographic Technique. The types and construction of electrodes used in electromyography vary widely The two main types of electrodes used for the study of muscle dynamics are surface (or "skin") electrodes and inserted (usually wire or needle) electrodes (described in detail by Basmajian, 1967)

Basically, an electromyograph is a high-gain amplifier with selectivity for frequencies in the range of 10 to several 1000 cycles per second (c p s) It has been suggested that the sharply-peaked spectra of motor-unit potentials derived with surface electrodes render the use of amplifiers with limited frequency-response quite practical, and it has been found that 20 to 200 c p s is the most desirable frequency range for the elimination of amplifier "noise," general nonmuscular "tissue noise" and movement artifact without significant loss of motor unit potentials

Function of Whole Skeletal Muscle

GROSS ORGANIZATION

Approximately 434 skeletal muscles, composed of some 250 million muscle fibers, account for 40 to 45% of the human body weight The muscle fibers are enmeshed in collagenous and reticular connective tissue At the ends of the elongated muscle fibers the connective tissue forms a common bundle of fibers, the tendon The connective tissue surrounding individual muscle fibers is called the endomysium Groups of muscle fibers, collected into bundles or fascicles are bound together by a more dense layer of collagenous and elastic fibers called the perimysium Finally the collagenous tissue which binds the fascicles into definitive muscle is called the epimysium The connective tissues of all these three layers are actually in continuity with each other

The arrangement of the muscle fibers in a muscle is variable (see Fig 1 97) and has been classified as follows

1) Muscle fibers parallel to the long axis of the muscle (e g , the strap like sartorius the pectineus, and the fusiform biceps brachii and flexor carpi radialis)

2) Muscle fibers oblique to the long axis of the muscle Muscles whose fibers are arranged so that they insert obliquely into a tendon in-

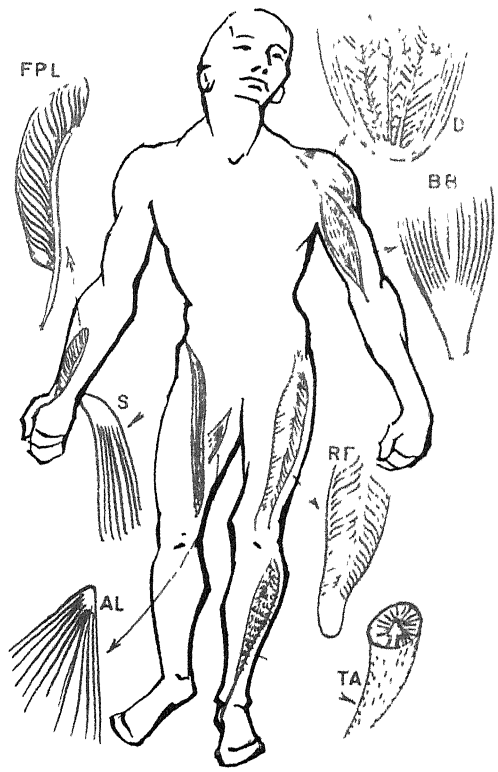


Fig 1 97 The form and distribution of several varieties of striated skeletal muscle The pennate, flexor pollicis longus (*FPL*), multipennate, deltoid (*D*), fusiform, biceps brachii (*BB*), bipennate, rectus femoris (*RF*) circumpennate tibialis anterior (*TA*) radial, adductor longus (*AL*) and strap-like sartorius (*S*) (The male figure redrawn from the First Muscle Tabula in *Fabrica* by Vesalius)

clude several varieties Because of their resemblance to feathers, with *a*) barbs unipennate (e g , flexor pollicis longus), *b*) bipennate (e g , rectus femoris), *c*) circumpennate (e g , tibialis anterior), and *d*) multipennate (e g , deltoid)

3) Muscle fibers arranged in a radial or triangular shape (e g , adductor longus, pectoralis minor)

Muscles whose component fibers are in parallel array are able to shorten over greater distances than muscles whose fibers are organized in the pennate manner The penniform arrangement of muscle fibers is associated with those muscles which exert great force rapidly over a short distance

In order for muscles to perform their function they are attached, for the most part, from one bone across a joint or joints to another bone by *tendons* or flattened sheets of connective tissue called *aponeuroses* The tendons and aponeuroses are formed of collagen, which is flexible but practically inextensible

The points of muscular attachment are called the origin and the insertion The origin of a muscle is usually the fixed or proximal attachment, while the insertion is the distal attachment to the bone which is moved There are exceptions and in many cases the designation of the origin and insertion of a muscle has become a matter of anatomical convention The force developed by an uncontrolled maximal contraction of a whole muscle is of little practical use in the movements of man and animal Therefore, the force and movement generated in a given circumstance by a given muscle depends upon the time-integrated individual activities of its component motor units For fine control of movement one or more motor units are employed As more force is required the number of stimuli carried by each motor unit nerve is increased, along with the recruitment of additional motor units The grading of muscular activity is the result, therefore, of asynchronous firing of the motor units of a whole muscle

In considering the function of muscles in man, it should be noted that no muscle acts alone even in the simplest movement, i e , a variety of muscles which are described as synergists and antagonists are involved in each action The synergists are muscles which act together to produce a movement which no muscle could produce alone An example of synergistic function can be seen in the adduction of the hand at the wrist The flexor carpi ulnaris will both flex and adduct the hand To produce adduction alone, the extensor carpi ulnaris must be brought into action to offset flexion and yet permit adduction of the hand by the flexor carpi ulnaris Antagonists, on the other hand, are muscles with opposed function For example, the biceps brachii and triceps brachii are mutually antagonistic with respect to movement at the elbow joint The muscle which produces the movement is termed the agonist while the muscle which opposes the movement is termed the antagonist The antagonist is usually reflexly inhibited when the agonist is brought into action, e g , if one holds a weight on the palm of the hand with the elbow flexed perpendicular to the body, the biceps brachii (in this case the agonist) develops tension while the triceps brachii (the antagonist) is flaccid This action is reversed—the triceps being contracted and the biceps relaxed—when the hand presses down onto a table with the arms flexed When however, the arm is extended and the elbow joint is fixed, both the biceps and the triceps are contracted

Not all striated muscles are associated with the movement of skeletal parts at joints or with the fixation or maintenance of joint stability as related to posture Muscles serve as sphincters when they encircle an orifice (e g , eyelid, lips, and anus) Striated muscle is found in the upper two-thirds of the esophagus in a tubular arrangement which aids in the swallowing mechanism The diaphragm is a thin muscular sheet which upon contraction serves to enlarge the thorax and compress the abdomen, effects which are associated with inspiration This action is reversed during expiration when the muscle relaxes In addition, there are certain cutaneous muscles which have at least one of their attachments to the skin, e g , the mimetic facial muscles and the panniculus carnosus of the trunk

TRANSMISSION TO THE SKELETON OF MUSCLE FORCE

The force generated by a muscle is transmitted to the skeleton by means of its tendons The submicroscopic collagen filaments of tendons extend along the surface of the individual muscle fibers at the tapering ends of the muscle fibers The surface area of the muscle fibers is fairly large here due to longitudinal folds of the

plasma membrane There is close contact between collagen filaments and the muscle fiber surface

The shape of a tendon varies with the shape of the muscle to which it is attached, but the muscle fibers are always oriented at an angle to the main direction of the tendon Muscles, in which the angle is relatively large, are called pennate muscles This angle, which never exceeds 10 to 20°, serves to prevent fraying of the tendon when the diameter of the fibers increases in connection with contraction and shortening

A tendon may form one or several flat sheets at the region of contact with the muscle, and the muscle fibers can impinge on one or both sides of such a sheet The length of the fibers extending between tendons is usually considerably shorter than the overall length of the muscle The distance bridged by the muscle fibers is related to the range of shortening under physiological conditions and to the fact that under these conditions the maximal shortening of individual muscle fibers is kept within a certain limit, viz , about 30% of the maximal length of the muscle in the body This is presumably an adjustment to the relationship between the maximal tension developed by a muscle as a function of its length

In places where a large force is required but where the range of shortening is limited, muscles are characterized by multipennate structure with the tendons forming several septa arranged in two planes In this case the effective cross-section of the muscle is considerably larger than the anatomical cross-section The septa of the tendons in the human deltoid muscle were found to develop gradually during embryologic growth and to be reduced in "old age " It is therefore likely that the gross structure of a muscle is dynamically adjusted to its functional requirements as is its cross-sectional area.

The cross-sectional area of a muscle increases when the muscle is forced repeatedly to develop maximal or close-to-maximal tension, as in regular exercise or work that requires forceful muscle action This increase is due to an increase of the cross-sectional area of the individual muscle fibers (hypertrophy) but not to a formation of new fibers (hyperplasia) Hypertrophy will persist and increase only if the work load is continuously increased during a training period If the load is kept constant the growth of the muscle stops when its strength has been adjusted to the load

Muscles exposed to sustained rhythmic activity over fairly long periods of time at a load below maximum show an increase in the density of the network of blood capillaries extending between the muscle fibers This appears as an adjustment to the requirements for increased oxygenation associated with this form of activity In animals experiments have shown the effect to be more pronounced in the young than in adults

The subdivision of the muscle tissue into individual muscles is, as was pointed out above, determined by mechanical factors, the muscle tissue being split up into individual muscles at places where large relative movements between parts of the muscle mass occur during movement of the body Such translational movements are most extensive between muscles that exhibit antagonistic effects, such as flexors and extensors Such muscles are also frequently separated by extensive connective tissue septa The individual muscles can therefore be classified with respect to their mechanical effects on the skeleton

The participation of a muscle in a particular movement depends on the pattern of impulses transmitted to it from the central nervous system Surprisingly these patterns can at times involve the activation of mechanically antagonistic muscles (or parts of muscles) and the failure to activate mechanically synergistic muscles It was found, for instance, by electromyographic recording, that the ventral and dorsal portions of the human deltoid muscle participate in the abduction of the arm during the whole range of the movement in spite of the fact that these components of the deltoid are mechanically antagonistic during part of this movement On the other hand, these components are not necessarily activated in adduction —even when the arm is so positioned as to render them mechanically synergistic.

Therefore, one cannot always deduce the patterns of muscle activity associated with a particular movement A direct recording of the muscle activity is required to give precise information regarding the patterns of activation of the various muscles participating in any particular type of movement. This consideration applies particularly to movement in ball joints

CORRELATED STRUCTURE AND FUNCTION OF SKELETAL MUSCLE FIBERS

It has been recognized for many years that muscles vary in color, not only from species to species but also within the same animal Striking examples of this are the crimson red pectoralis of the pigeon as contrasted with the stark white chicken pectoralis and the red soleus lying near the white gastrocnemius in the lower limb of animals and man In birds (falcon, goose, and chicken) it has been shown that the amount of red pigmentation in the pectoralis muscle can be correlated directly with the capacity for sustained flight Furthermore, it was shown experimentally that it takes fewer stimuli per second to produce a smooth, sustained, maximal contraction (tetanus) in the red muscles than in the white muscles A generalization derived from many investigations indicates that all slow-contracting muscles have red pigmentation, but not all red muscles are slow-contracting (e g , the soleus is a red, slow-contracting muscle, while the masseter is also a red muscle but it contracts more rapidly)

In more quantitative terms, whole muscles provide a spectrum of speeds of contraction and relaxation as shown in Figure 1 98 It can be seen that the contraction time of an extraocular muscle (7 5 msec at 37° C) is 5 times faster than that of the gastrocnemius muscle (40 msec) and 12 times faster than that of the soleus muscle (90 msec) The extraocular muscles and the soleus represent the fast and slow extremes in most mammals It is important to remember that whole muscles are not normally

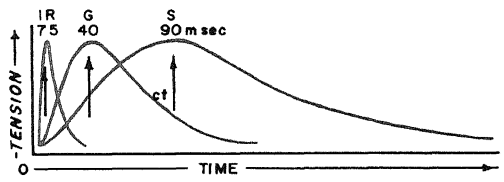


Fig 1 98 Illustration of the isometric contraction-relaxation (twitch) curves for three mammalian skeletal muscles The lateral rectus (*L R*) gastrocnemius (*G*) and soleus (*S*) represent the extremes (fast and slow) and midrange in a spectrum of contraction-relaxation times for mammalian muscles The arrows indicate peak tension development which corresponds to the contraction time for each muscle (lateral rectus 7 5 msec gastrocnemius 40 msec , and soleus 90 msec)

activated synchronously to yield the maximal twitch response of the kind recorded in Figure 1 98 Furthermore, the twitch response of isolated whole muscles does not reveal the heterogeneous nature of the tissue either in functional or structural terms

Microscopic examination of histological sections of muscle reveals variability in cross-sectional diameter, internal structure, and certain histochemical reactions of the muscle fibers An examination of the structural-functional components of the gastrocnemius also illustrates the heterogeneity of muscle The white gastrocnemius muscle of the cat is in the midrange of muscle speed (contraction time is about 40 msec and contraction-relaxation time about 160 msec) The contraction time of the motor units which comprise the gastrocnemius vary from 17 8 to 129 msec with a distribution peak between 30 and 40 msec The peak distribution corresponds to the contraction time for the twitch of the isolated whole muscle In addition, not all motor units produce a twitch response, however, the slow-responding motor units do develop tension when given multiple (tetanic) stimulation

In general, large motor units have a higher threshold, contract more rapidly, and develop a greater degree of tension than the smaller motor units

Histochemical and electron microscopic studies reveal that the gastrocnemius is composed of three distinct types of muscle fibers About 50 to 55% of the muscle fibers and 70% of the cross-sectional area are composed of fibers which are large and contain few mitochondria, as demonstrated by histochemical localization of succinic dehydrogenase and adenosine triphosphatase activity About 30% of the muscle is composed of smaller dark cells with a greater density of mitochondria The remaining 14 to 20% is composed of the smallest dark fibers with densely packed mitochondria The large fibers are the classical white, fast twitch fibers, while the smaller dark fibers represent two types of red, slow contracting fibers The soleus muscle of the cat (but not those of the rabbit and rat) is composed entirely of the same intermediate dark fibers found to comprise 30% of the gastrocnemius Therefore, these fibers can be classified as *slow twitch* muscle fibers The smallest dark fibers in the gastrocnemius are probably *slow* muscle fibers,

judging from the physiological characteristics of some of the motor units in the gastrocnemius

Although the cat soleus muscle appears to be structurally homogeneous its motor unit contraction time varies from 50 to 193 msec The peak distribution of motor units has a contraction time of 80 to 90 msec , which corresponds to the contraction of the twitch of an isolated whole muscle Approximately 90% of the soleus motor units develop 70% of their maximal tension at low frequency stimulation i e 20/sec Some units develop a large fraction of their contractile power at 5 to 10 stimuli/sec The soleus can maintain tone or tension as required by slow, low frequency firing of its various motor units Experimental evidence also indicates that there is an overlap in contraction times (from 50 to 129 msec) between the slower motor units of the gastrocnemius and the faster motor units of the soleus muscle

Even the fast contracting extraocular muscles from the cat are composed of three morphologically distinct muscle fibers The isolated muscle produces both a twitch-type contraction after single shocks and a slow, graded nontwitch contraction which is elicited only after multiple stimulation The variety of structural and physiological properties of these muscles are, therefore, basically similar to those of the slower gastrocnemius

THE INNERVATION AND SUBMICROSCOPIC ORGANIZATION OF TWITCH AND SLOW MUSCLE FIBERS

The twitch muscle fibers can be characterized physiologically by their ability to produce (on nerve stimulation) a propagated action potential and to respond with fast shortening or tension development Within this group many differences exist in color, contractile behavior, rates of conduction and contraction time The slow fibers, however, differ fundamentally from the twitch fibers because they have lower levels of resting potential (50 to 70 mv as compared to 80 to 90 mv for fast fibers), they do not produce propagated action potentials, they exhibit a longer latency of response (6 to 8 msec as compared to 2 to 4 msec for fast fibers)—and they do not respond to stimulation by rapid shortening or rapid tension development Only repeated stimulation results in a significant rise in tension by these slow muscle fibers In addition the slow muscle fibers are innervated by numerous nerve endings (the en grappe type), distributed widely over the whole fiber length, whereas the twitch fibers are innervated by a single discrete localized nerve ending (the en plaque type) with a distinct neuromuscular junction The numerous areas of nerve contact which locally activate the contractile apparatus of the fiber compensate for the inability of a slow fiber to produce propagated muscle impulses

Although the slow muscle fiber system is widely found it has been studied in greatest detail in the frog In this animal tension development by slow fibers following reflex activation is similar to tension development after direct nerve stimulation (provided that the direct stimulation is of sufficient intensity to activate most of the small nerve fibers serving the muscle) The motor units of the slow fiber system may be normally activated, therefore, by a synchronous high frequency discharge from the spinal cord In the frog the nerve fibers to the slow muscle fibers are small and slow conducting (2 to 8 m /sec) while the nerve fibers to the twitch fibers are large and fast conducting (8 to 40 m /sec) No overlap of innervation between the two muscle groups has been found The slow muscle fibers are reflexly activated even during periods when the twitch fibers are at rest The slow fibers can also be activated during twitch fiber activity and can maintain large tensions originally produced by the twitch fibers over very long periods by virtue of low frequency stimulation A functionally useful synergism has been demonstrated by the twitch and slow fibers in the frog tibialis muscle First, small nerve stimulation at 30/sec results in a smooth rise in tension If a burst of twitch activity is added, via the large nerve system, the final tension developed exceeds that which would have been developed without the twitch activity Secondly, when stimulation ceases there is a very gradual relaxation of the generated tension If a burst of twitch activity is initiated during the slow relaxation phase, the residual tension in the slow fibers collapses rapidly

Slow fibers and twitch fibers have similarities and differences in their fine structure In both, the myofilaments of the myofibrils are of the same length even though sarcomere lengths may differ However, the slow fibers have no discrete M lines, the H zones are poorly de-

fined and the Z lines are thick Although the myofibrils tend to be large and vaguely defined in cross-section, this cannot be considered as a sole criterion for slow fibers

A striking difference between twitch and slow fibers is the absence or poor development of the transverse tubular or T system in the slow fiber In the twitch fiber the T tubule is in continuity with the sarcolemma, but, in those slow fibers in which a T system can be recognized, the small delicate tubules have no specific localization with respect to the sarcolemma or the contractile apparatus Thus, local contractions are not confined to specific areas of the myofibril The smallest contractile unit of many twitch fibers is that specific area of the myofibril between two adjacent M lines, the T-system being located at the Z line In other twitch fibers, the smallest contractile unit is the half sarcomere and, in this case, the T system is found at the level of the junction of the A and I bands (see Fig 1 99)

The motor nerve endings on slow muscle fibers, as revealed by electron microscopic studies, are readily distinguishable from those of twitch muscle fibers The specific region of nerve-muscle contact is small compared with that on twitch fibers In addition, there are no junctional folds in the muscle membrane or sarcolemma in the region of the nerve ending Such junctional folds are a constant feature of the motor end-plate in twitch fibers

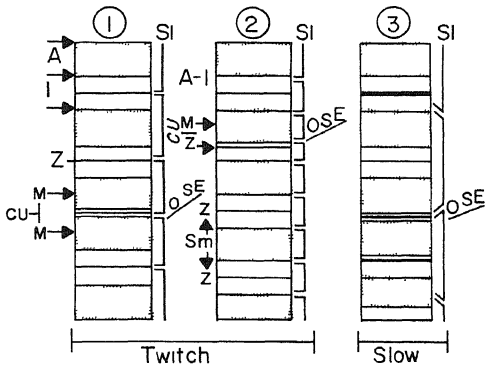


Fig 1 99 Local excitation of twitch and slow muscle fibers The response to local subthreshold stimulation (*SE*) delivered to the sarcolemma (*SI*) is specific and discrete in twitch fibers In twitch fiber 1 a localized contraction can be elicited in the region of the Z line This is the precise localization of the transverse tubules or T system in this fiber The smallest *contractile unit* (*CU*) therefore is that area of the myofibril between two adjacent M lines (M) *A* A band *I* I band *Z* Z line

In fiber 2 the local response is found at the level of the A band-I band junction (*A I*) The T system in this fiber is at the A band-I band junction The *contractile unit* (*CU*) is thus a half sarcomere or that area of the myofibril between adjacent Z and M lines The sarcomere (*SM*) includes one A band and two half I bands Fiber 3 represents a slow muscle fiber Stimulation of this fiber results in a variable or diffuse response No portion of the fiber is selectively excitable The T system when found in slow fibers has a meandering course

Among the various factors that may account for the difference in the slow and fast fibers, the nerve endings appear to have the most influence For several years, cross innervations have been made between slow and fast mammalian muscles The procedure is to cut the motor nerve to these fibers and cross them so that slow muscle fibers are innervated by fast firing nerves and fast muscle fibers are innervated by slow firing nerve endings It has been shown that after recovery from the cross innervation procedure, the slow fibers have a modified behavior They contract more rapidly and they undergo morphological and biochemical alterations, e g , methyl histidine appears in myosin and there is an enhancement of the activities of several enzymes inducing the ATPase associated with myosin The main effect of cross innervation of fast muscle with slow nerve fibers is a slowing of the twitch

It has been suggested that changes in the rate of firing of nerve fibers innervating skeletal muscle fibers may be a factor in the differentiation of slow and fast fibers There are some indications that if a fast muscle is cross-innervated with slow nerve fibers, but electrically stimulated at a rate of 30 to 60/min (normal fast fiber rate), a change from fast muscle properties to slow muscle properties is prevented Although we have very limited information concerning the exact role of motor nerves as a trophic influence, it is well-established that the integrity of the nerve supply is of utmost importance for the preservation of normal muscle structure and function, e g , sectioning a motor nerve is always followed by atrophy as well as paralysis of the muscle to which it is connected It is highly probable therefore that a trophic factor is synthesized in, or at least conveyed to, the muscle through its nerve

DIFFERENTIATION OF FAST AND SLOW TWITCH MUSCLES

The growth and development of mammalian muscles with different contraction times has

been given much attention In general, all muscles in newborn animals are equally slow It takes about 6 weeks (in cats) for the fast twitch fibers to attain their adult speed The more slowly contracting muscles, such as the soleus and crureus, pass through a short period during which their contraction times quicken, but this is followed by a progressive slowing until the adult condition appears in about 20 weeks It can be shown that the differentiation of the slow contraction time can be abolished if the cord is transected or if the lumbosacral spinal cord is isolated from incoming stimuli In this case, the muscles do not become slow, but after a few weeks the soleus and crureus become almost as fast as normal fast muscles (e g , gastrocnemius) In adult rabbits, a similar transformation in the soleus will occur if the tendinous insertion of the muscle is cut (tenotomy) Electromyographic studies show that the soleus is normally continuously activated Upon tenotomy, the continuous activity ceases even with intact motor innervation It can be concluded that motor impulse activity plays a fundamental role in determining the speed of contraction in mammalian skeletal muscle, although the exact mechanism is not known

Following tenotomy, the red, slow contracting soleus is transformed into a white, fast contracting muscle Not only have the mechanical properties of the soleus been changed, but the red color has disappeared The red color of muscle is due to myoglobin which is related to the hemoglobin of red blood cells Myoglobin from skeletal muscles has a molecular weight of 16,700 and contains 1 heme (iron-porphyrin complex) per molecule The hemoglobin of the blood is composed of 4 hemes/molecule and differs also in the protein globin Myoglobin has a greater affinity for oxygen than does hemoglobin, and it serves as an oxygen reservoir within the muscle fiber In addition to its high affinity for oxygen, myoglobin loads and unloads its oxygen with great speed Myoglobin is probably synthesized within the muscle fiber much as hemoglobin is synthesized within the red blood cell

A direct relationship has been found between the myoglobin content of muscle and the rate of blood flowing through a given muscle mass, usually referred to as the fractional blood flow and expressed as the per cent of cardiac output perfusing 100 g of muscle

The density of the capillary network (as measured by the alkaline phosphatase reaction) is highest in the red fibers lowest in white fibers, and intermediate in "intermediate" fibers Myocardial fibers and the external oculomotor fibers have a denser capillary network than other muscles Twitch time is proportional to the fractional flow (thus, red muscles twitch more slowly than white muscles)

The myosin ATPase activity is greatest in fast muscle fibers least in slow fibers, and intermediate in "intermediate" fibers The pH optimum is in the alkaline range for the ATPase of white (fast) fibers and in the acid range for the ATPase activity of slow fibers, the "intermediate" fibers have maximal ATPase activity in an intermediate pH range In addition to ATPase, the properties of aldolase and other enzymes have also been found to differ when red and white fibers were compared The high myosin ATPase, which probably is related to the ability to shorten rapidly, serves to provide the speed and acceleration needed in the functions that the white muscles serve

Metabolism of the red fibers is mostly aerobic while white muscles can also function anaerobically as illustrated by the increase in the lactic acid content in the veins draining an exercising muscle Such anaerobic metabolism leads to "oxygen debt," from which the muscle must ultimately recover by oxidative replenishment of its energy stores During fetal development while white fibers are differentiating, the histidine and lysine moieties of the muscle proteins are methylated It is estimated that white fibers contain 1 mole of 3-methyl-histidine per mole of myosin, whereas fetal, red, and cardiac muscle fibers do not contain any 3-methyl histidine The ATPase activity of the fibers increase three fold with the appearance of 3-methyl histidine. Actin fibers also contain 3-methyl histidine, however, the methylation in actin is unrelated to that in myosin

Red slow muscle fibers undergo sustained rhythmic activity They respond to stimuli with prolonged contractures. They constantly adjust to postural change They operate during periods of prolonged anoxic stress as demonstrated by diving mammals Accordingly, those animals whose activity is sustained generally have red muscles The lower myosin ATPase in red muscle permits muscular contraction to proceed with slower utilization of ATP This

facilitates the appropriate balance between chemical energy supply and mechanical energy utilization that must occur in a muscle that contracts steadily and without rest for long periods of time.

White fast muscles undergo only short bursts of activity and possess little myoglobin. It should be recalled, however, that most muscles are a mixture of both red and white fibers. Nature has devised a mechanism, whereby the neuromuscular system utilizes both tonic and phasic contractions in order to provide an exquisite range of functional activity about which there is still much to learn.

5 / AMINO ACIDS AND PROTEINS

The Molecular Framework and Machinery of Living Systems

Proteins are the realization of the information contained in deoxyribonucleic acid (DNA) Through them this information is transformed into physiological work They are the primary constituents of cellular machinery and form an integral part of the regulatory systems that determine how the machinery operates

The ubiquitous distribution of the various classes of proteins reflects the roles they serve in the body

1) As *structural elements* they are found in the membranes of cells and organelles They are a major component of the various fibers of connective tissue—collagenous, reticular and elastic—as well as the ground substance

2) Their function as *enzymes* is legion As such they are responsible for the ordered sequential reactions which result in chemical and biological work This makes the thermodynamically unfavorable process called "life" possible Special classes of enzyme proteins fulfill a mechanochemical role Actin, myosin, and the related muscle proteins are involved in the conversion of chemical energy [adenosine triphosphate (ATP)] into mechanical work

3) *Intramembrane carrier proteins* act as shuttles, mediating the transport of various molecules and ions across the cell membrane When energy is required for the passage of substances across the cell membrane (active transport) it is the carriers and their allied cofactors that generate a vectorial quality for the net movement of solutes into or out of the cell

4) Proteins also serve a *carrier function in plasma* by binding small molecules and ions and transporting them throughout the body Thus, thyroxin-binding globulin transports thyroxine to cells where it is released and rendered active Transferrin, an iron-binding globulin, delivers Fe^{++} to hemoglobin-synthesizing loci The activity of ions such as Ca^{++} can be controlled by the ratio of the free to the bound species When H^+ ion is bound, the protein acts as a buffer to minimize changes in pH Since cells are impermeable to proteins, they also participate in determining the ionic distribution and, hence the electrical potential difference across cell membranes (Gibbs-Donnan equilibrium) By virtue of its osmotic activity albumin mediates the distribution of body fluid between the plasma and extracellular compartments

5) The *antibody proteins* (immunoglobulins) contribute to the body's defense against infection

6) Acting as *hormonal messengers*, peptides and proteins (e g , insulin, growth hormone, vasopressin, etc) carry information that regulates cell permeability and cell metabolism

7) They provide a *source of fixed nitrogen* for the synthesis of essential nonprotein nitrogenous compounds such as heme, creatinine, etc

8) After hydrolysis, the constituent amino acids can be deaminated and metabolized to CO_2 and H_2O, and thus serve as *sources of energy*

One of the most important properties of living systems, that of specificity, is expressed through proteins This property is determined by the primary protein structure as dictated by the nucleic acid sequence in DNA and, in turn, by the conformational consequence of this primary structure Each animal species and each tissue within an organism can be characterized by its component proteins Indeed, it is the variation of protein content among tissues which results in functional specialization and tissue specificity (differentiation) This arises from the fact that each tissue selects amino acids in a particular ratio from the total body pool of amino acids and polymerizes them into proteins characteristic for that tissue The varieties of proteins so produced (Table 1 7) were classified at the turn of the 20th century chiefly on the basis of solubility properties, and there seems to be little prospect of a better classification scheme in the near future

What is common to all proteins is the fact

Table 1 7

Class of Protein	Characteristics	Examples
	A Simple proteins	
1) *Albumins*	Soluble in water and coagulable by heat Present in both animal and plant tissues	*Serum albumin, egg albumin, lactalbumin*, and various vegetable albumins such as *leucosin* (in wheat, rye, and barley), *legumelin* in lentils, soy beans, beans and peas, and *phaselin* in kidney beans
2) *Globulins*	Soluble in dilute saline solutions, insoluble in water Animal globulins are coagulated by heat Vegetable globulins imperfectly or not coagulated by heat	*Serum globulin, fibrinogen* (and *fibrin*) *vitellin* of egg yolk, and vegetable globulins such as *excelsin* (Brazil nuts), *edestin* (hemp), *phaseolin* (kidney beans), *legumin* (peas and lentils), and *tuberin* (potatoes) A number of other vegetable globulins have been isolated and named
3) *Glutelins*	Found only in plants Insoluble in water, saline, or alcohol, but soluble in very dilute alkali	*Glutenin* of wheat, *oryzenin* of rice, and *glutelin* of maize
4) *Prolamines or gliadins*	Found in cereals (except rice) Soluble in 70 to 90% alcohol, insoluble in water They contain a large proportion of proline and compounds giving rise to ammonia-nitrogen	*Gliadin* of wheat, *hordein* of barley, and *zein* of maize
5) *Albuminoids or scleroproteins*	Especially resistant to the usual reagents They enter into the construction of protective and connective tissues, e g , skin, tendons, ligments, and bones	*Keratin* of hair, skin, bone, feathers, tortoise shell and eggshell, *elastin, collagen, ossein* and *gelatin* of tendons, ligaments, bone, etc
6) *Histones*	Soluble in water and precipitated by ammonia solution and by alkaloids They contain a large percentage of basic (diamino) acids (Table 1 8)	*Globin* of hemoglobin, *thymus histone, scombron* and *gadus histone* in spermatozoa of mackerel and codfish, respectively
7) *Protamines*	Found in combination with nucleic acid in heads of fish spermatozoa Constructed predominantly of basic amino acids	*Salmine* and *sturine* in spermatozoa of salmon and sturgeon, respectively
	B Conjugated proteins Proteins whose molecule is combined with another nonprotein group	
1) *Nucleoproteins*	Nucleic acid in combination with a protein belonging usually to the class of histones or protamines	Present in cell nuclei
2) *Chromoproteins*	Protein in combination with a pigment (e g , hematin) containing iron, copper, or other metal	Hemoglobin, hemocyanin, etc
3) *Glycoproteins*	Proteins other than nucleoproteins in combination with a carbohydrate group	*Mucin* in salivary gastric and intestinal secretions, *ovomucoid* of egg white and *chondromucoid* of cartilage
4) *Lipoproteins*	Proteins in combination with lipid	Present in plasma, milk, cell nuclei
5) *Phosphoproteins*	Proteins other than nucleoproteins and lecithoproteins in combination with a phosphorus-containing group	*Caseinogen* (and casein), *vitellin* of egg yolk

C *Derivatives of proteins—derived proteins* These are produced by the action of acids, alkalis or proteolytic enzymes upon certain of the proteins listed above		
a) *Primary derivatives*		
1) *Proteans*	Insoluble products formed in the early stage of the action upon proteins of water, dilute acids, and enzymes	
2) *Metaproteins*	Formed in a later stage of the action of acid or alkali	*Acid* metaprotein, *alkali* metaprotein
3) *Coagulated proteins*	Formed by the action of heat or of alcohol upon solutions of proteins	
b) *Secondary derivatives*		
1) *Proteoses*	Formed by the action of pepsin or trypsin upon proteins. They are soluble in water from which they are precipitated by saturation with ammonium sulfate. They are incoagulable by heat	*Albumose* from albumen, *globulose* from globulin, *caseose* from casein
2) *Peptones*	These represent a further stage in action of proteolytic enzymes. They are soluble in water but are not precipitated from an aqueous solution by ammonium sulfate. They are not coagulated by heat	
3) *Peptides, dipeptides, tripeptides* and *polypeptides*	Products formed in the final stages of proteolytic digestion	Glycyl-alanine, leucyl glutamic acid, etc

that they are made up of amino acids (Table 1-8). What is specific about proteins is the ratio of the component amino acids and the order (primary sequence) in which they are polymerized. Therefore, a consideration of the amino acids is the logical starting point for a survey of the physiology of proteins.

Physicochemical Properties of Amino Acids

Amino acids are the monomeric building blocks which, as noted above, are joined together in a specific sequence to form proteins. The type and configuration of amino acids within the plant and animal kingdoms are not in all cases identical. For instance, within the plant kingdom, amino acids can be found that occur in the *D* as well as the *L* configuration (as in bacterial cell walls) or have various modifications in the side chain (e.g., hydrogen replaced by a halogen, etc.). There are 24 naturally occurring amino acids found in protein hydrolysates from animal tissues (Table 1-7). Other amino acids are not polymerized into proteins but found as monomers in tissue, e.g., γ amino butyric acid (GABA). The general structure for an amino acid is as shown in Figure 1-100. The penultimate carbon atom, designated α, is covalently bonded to an amine or imine function, a carboxylic acid function and an R group which differentiates one amino acid from another.

Thus, fundamentally, it is the sequence of R groups of a chain of amino acids which accounts for the characteristic structural and functional properties of proteins.

The R group is simply H in the case of glycine but it may be an aliphatic, aromatic or heterocyclic group in other amino acids. Within these larger categories the R group is further delineated by a variety of specific substituents (Table 1-8).

ACID-BASE PROPERTIES OF AMINO ACIDS

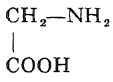
In aqueous solution amino acids and proteins behave as weak acids or weak bases depending

Table 1 8
Classification of the amino acids[a]

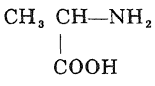
I Aliphatic Amino Acids

A Monoamino-monocarboxylic acids

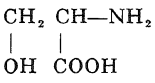
1) Glycine (or glycocoll) $C_2H_5NO_2$, or amino-acetic acid



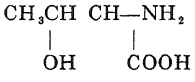
2) Alanine, $C_3H_7NO_2$, or α-amino-propionic acid



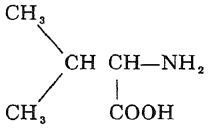
3) Serine, $C_3H_7NO_3$, or α-amino-β-hydroxy propionic acid



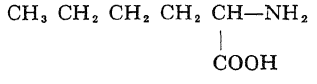
4) Threonine, $C_4H_9NO_3$, or α-amino-β-hydroxy-*n*-butyric acid



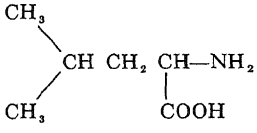
5) Valine, $C_5H_{11}NO_2$, or α-amino-isovaleric acid



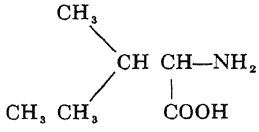
6) Norleucine, $C_6H_{13}NO_2$, or α-amino-*n*-caproic acid



7) Leucine, $C_6H_{13}No_2$, or α-amino-isocaproic acid

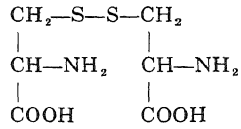


8) Isoleucine, $C_6H_{13}NO_2$, or α-amino-β-ethyl-β-methyl-propionic acid

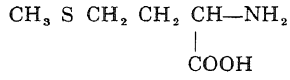


Sulfur-containing monoamino-monocarboxylic acids

9) Cystine, $C_6H_{12}N_2S_2O_4$, (or dicysteine) or di-(β-thio-α-amino-propionic) acid

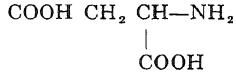


10) Methionine, $C_5H_{11}SNO_2$, or α-amino-γ-methyl-thiol-*n*-butyric acid

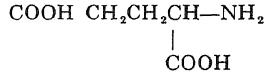


B Monoamino-dicarboxylic acids

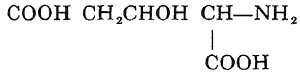
11) Aspartic acid, $C_4H_7NO_4$, or α-amino-succinic acid



12) Glutamic acid, $C_5H_9NO_4$, or α-amino-glutaric acid

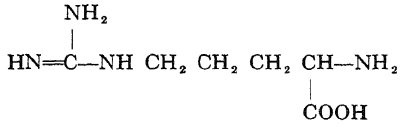


13) Hydroxyglutamic acid, $C_5H_9NO_5$, or α-amino-β hydroxy-glutaric acid

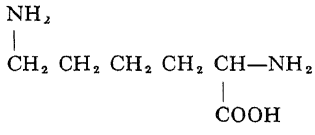


C Diamino-monocarboxylic acids

14) Arginine, $C_6H_{14}N_4O_2$, or α-amino-δ-guanidine-*n* valeric acid

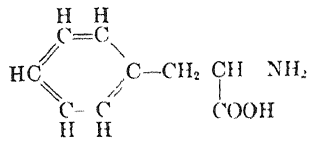


15) Lysine, $C_6H_{14}N_2O_2$, or α-ϵ-*diamino-caproic acid*

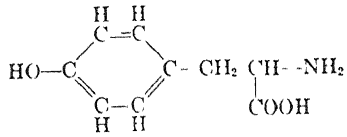


II AROMATIC AMINO ACIDS

16) Phenylalanine $C_9H_{11}NO_2$ or α-amino β-phenyl propionic acid

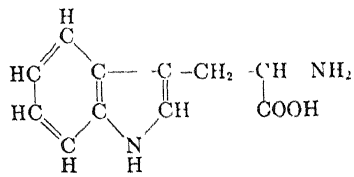


17) Tyrosine, $C_9H_{11}NO_3$, or α-amino-β-parahydroxy phenyl propionic acid



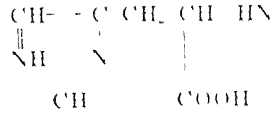
III HETEROCYCLIC AMINO ACIDS

18) Tryptophane $C_{11}H_{12}N_2O_2$, or α-amino-β-indole propionic acid



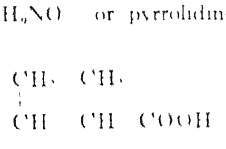
Indole nucleus

19) Histidine $C_6H_9N_3O_2$ or α-amino β-imidazol propionic acid



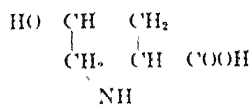
Imidazole ring

20) Proline $C_5H_9NO_2$ or pyrrolidine α-carboxylic acid



Pyrrole nucleus

21) Hydroxyproline (oxyproline) $C_5H_9NO_3$, or γ-hydroxy pyrrolidine α-carboxylic acid



[a] *Hydroxylysine, thyroxine, 3-5-duodotyrosine* or iodogorgoic acid, *citrulline (α-amino δ-carbamido n-valeric acid)* and *ornithine(α,δ-diamino-valeric acid)*, and several others would have to be added to complete the list of naturally occurring amino acids

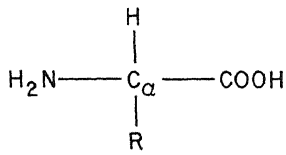


Fig 1-100 General structure of an amino acid

on the number and nature of acidic and basic R group substituents Thus, proteins and amino acids serve as physiological buffers in the extracellular and intracellular fluids Weak electrolytes (acids and bases) do not ionize readily Consequently, addition of a proton (H^+) or hydroxyl ion (OH^-) to such a system removes the ion from solution so that the pH tends to remain constant, as shown in Figures 1-101 and 1-102

Amino acids and proteins, thus behave as Bronsted acids and bases (proton donors and proton acceptors, respectively) As such, information concerning their pH behavior can be obtained by titration pH is defined as log $[H^+]$ where $[H^+]$ represents the molar concentration of hydrogen ions The acid dissociation constant, Ka, for the reaction, HA $H^+ + A^-$, is

$$[H^+][A^-]/[HA] \text{ at equilibrium}$$

Therefore

$$[H^+] = ([HA]/[A^-])\ Ka$$

Taking the negative logarithm of both sides of this equation yields

$$-\log [H^+] = -\log ([HA]/[A^-]) + (-\log Ka) \text{ or}$$
$$pH = pKa + \log [A^-]/[HA]$$

This is the Henderson-Hasselbach equation which relates the pH of a system to the degree of ionization of the species under consideration

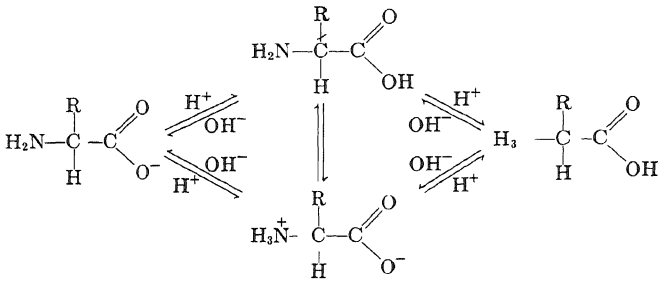


Fig 1 101 Titration of an amino acid

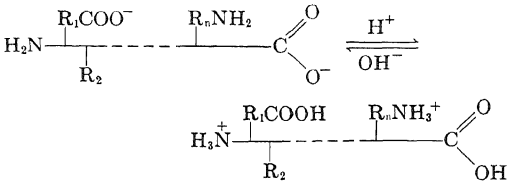


Fig 1 102 Titration of a protein

At 50% ionization the ratio $[A^-]/[HA]$ being 1 and log 1 being 0,

$$pH = pKa$$

Operationally the pKa of the ionizable groups of an amino acid or protein can be determined as the points of inflection observed by plotting the results of a titration of the amino acid or protein in terms of pH vs the amount of added base or acid Figure 1 103 illustrates the titration of the aromatic amino acid tyrosine starting with its fully protonated form (acidic solution) Since there are three groups involved in proton donation (NH_3^+, phenolic hydroxyl, carboxyl) there will be three stages in the titration, each requiring 1 Eq of base (NaOH) to reach the end point Each stage has a pKa which is characteristic of the group being titrated Since it requires a total of 3 Eq of NaOH to neutralize the three groups, tyrosine is an example of a polybasic acid All amino acids exhibit this amphoteric behavior Figure 1 104 illustrates the dissociation stages of glycine, a neutral amino acid, glutamic acid, an acidic amino acid, and lysine, a basic amino acid Examination of the various stages of dissociation of these and other amino acids reveals an H^+ ion concentration, characteristic for each amino acid, at which it has a net charge of zero, i e , the sum of its positive charges equals the sum of its negative charges This H^+ ion

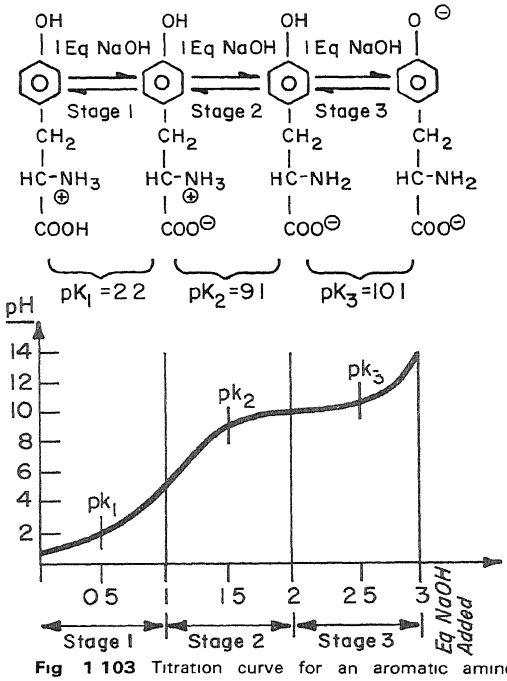


Fig 1 103 Titration curve for an aromatic amino acid tyrosine showing three stages of titration

concentration is the *isoelectric point* of the amino acid, the corresponding pH is designated by the symbol pI At its isoelectric point an amino acid behaves as an ampholyte (zwitterion), in other words, at its pI, an amino acid is electrically neutral and it can behave as an acid or a base For acidic amino acids pI < 7, for basic amino acids pI > 7 Most amino acids are zwitterions in the physiological pH range

Zwitterions are dipoles Schematically they may be represented as shown in Figure 1 105

There are several consequences of this dipolar structure which affect the physical properties of amino acids in solution and in the crystalline form and, indeed, the physical proper-

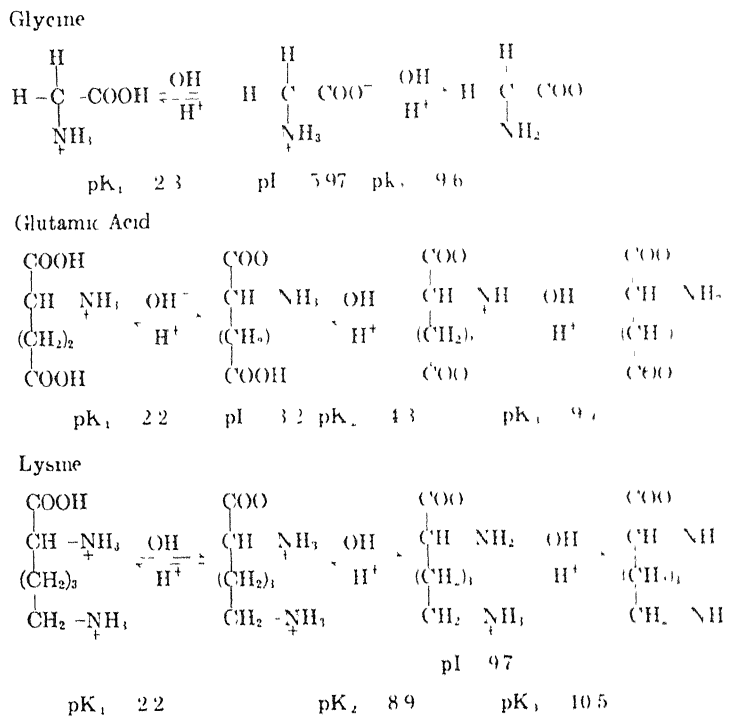


Fig 1 104 Dissociation stages of glycine (a neutral amino acid) glutamic acid (an acidic amino acid) and lysine (a basic amino acid)

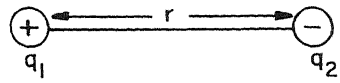


Fig 1 105 Structure of a zwitterion

ties of an amino acid confirm that it exists in the form of a zwitterion rather than in the form of a neutral or an uncharged molecule

A dipole consists of charges (q_1, q_2), opposite in signs but equal in magnitude, separated by a fixed distance (r) When a dipole is placed in a uniform, external electric field, it will spin about the center of the fixed distance in order that net force on it will be zero (Fig 1 106) The product of the charge, q, and the distance, r, is the dipole moment, and amino acids are found to have appreciable dipole moments as would be expected for zwitterions Inspection of the final state diagram indicates that since the charges are equal and opposite, the resultant force is zero. Consequently, the dipole experiences no net force and is stationary in an electric field, i e , it does not migrate to the cathode or anode Accordingly, at its pI, an amino acid does not migrate in an electric field

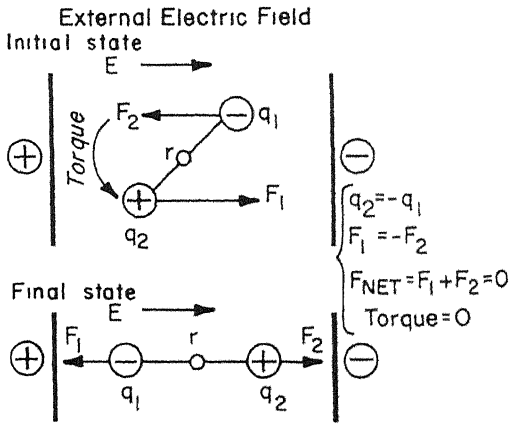


Fig 1 106 Diagram illustrating rotation of a dipole in presence of uniform external electric field

When two or more dipoles come together they orient themselves in a specific way such that they interact electrostatically as sketched in Figure 1 107 Thus, amino acid crystals melt with decomposition because additional energy must be supplied to break the intermolecular electrostatic bonds This extra energy is in excess of that required to reach the melting point and the crystal undergoes decomposition in-

stead of a simple transformation to the liquid state Amino acids are more soluble in polar solvents (e g , water) than in nonpolar organic solvents, and they raise the dielectric constant of any solution in which they are dissolved Spectrographic analysis also indicates that amino acids exist as dipoles When amino acids are joined to form proteins, the dipoles arrange themselves so that they cancel themselves out and leave the overall protein structure as a net dipole

Physicochemical Properties of Proteins

PEPTIDE BONDS

In the synthesis of the protein molecule chains of amino acids are formed by the linkage of the carboxyl group of one amino acid to the α amino group of another amino acid with the concomitant liberation of a molecule of water (*peptide bond* formation) This joining of two amino acids through a secondary amide linkage (peptide bond), viz , —C—N—, leads

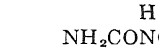


to the formation of a dipeptide (Fig 1 108) The similar addition of another amino acid will yield a tripeptide In this manner polypeptides are built up In representing peptides and proteins the N-terminus (amino-terminus) is written on the left and the C-terminus (carboxyl-terminus) on the right

There is a large body of evidence which indicates that in peptides and proteins, the amino acids are indeed linked via peptide bonds For example, the structure of myoglobin, an iron containing protein in muscle, has been determined by Kendrew (1962) and his associates, using x-ray analysis at a resolution of 1 5 A which allows individual atoms to be distinguished The electron density analysis clearly showed that the interamino acid linkages were peptide bonds

In the chemical synthesis of peptide and protein hormones such as oxytocin, vasopressin, insulin, etc the individual amino acids are linked together by classical methods for the formation of peptide bonds and the final products exhibit the same physicochemical properties and biological activities as their naturally occurring counterparts The peptide bond gives a characteristic infrared and far-ultraviolet absorption spectrum, natural and synthetic peptides and proteins give analogous spectra

The fact that proteins give an intensely positive biuret reaction provides a useful if less direct evidence for the peptide bond hypothesis Biuret, a compound with the following structure,



reacts with an alkaline solution of copper sulfate to form a purple colored complex, the intensity of which is related to the amount of biuret present

When an intact soluble protein is titrated over a broad pH range, it is found that relatively few amino and carboxyl groups are titratable These titratable groups are the terminal free amino and carboxyl functions of the protein and the acidic and basic groups of the side chains (R groups) of amino acids such as arginine, lysine, asparagine, glutamine, serine, tyrosine, histidine, etc If the protein is hydrolyzed, the number of titratable amino and carboxylic groups increase in a 1 1 ratio and at equal rates

The characteristics of the peptide bond have been analyzed by x-ray diffraction There are

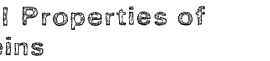


Fig 1 107 Electrostatic interaction of two dipoles

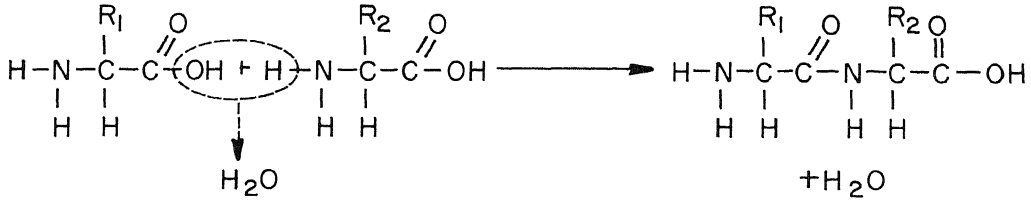


Fig 1 108 Formation of peptide bond linking two amino acids to yield a dipeptide

six atoms involved in this bond and generally it is coplanar (Fig 1 109) From the length of the C-N bond, it was determined that it is of 50% double bond character Thus, rotation about this bond is restricted and consequently the α carbons are trans to each other

Peptide linkages form the backbones of proteins Therefore, a generalized protein may be represented as in Figure 1 110

It will be seen that projecting off the backbone of this prototype protein are the various amino acid residues The primary structure of a protein is defined as the linear sequence of the peptide-bonded amino acids The projecting side chains (R groups) of the constituent amino acids are important in determining the superstructure of proteins

From the primary structure analysis of hemoglobin, myoglobin, carboxypeptidase A, lysozyme, ribonuclease (RNase) and others, it is clear that most amino acids occurring in proteins (from animal sources) are of the L-configuration and that the ordering of amino acids in the protein is random although genetically determined Given one amino acid out of the 20, there is an equal probability for any one of the other 20 amino acids to occur next to it, i e , any dipeptide sequence is as likely as any other

Considering a generalized protein (such as in Fig 1 100, where the nature of the R groups are left unspecified) there is sufficient flexibility in the covalently bonded peptide backbone to allow this polymer to assume a large number of conformations The conformation of the backbone is termed the secondary structure of the protein Before discussing the theoretically predicted and experimentally confirmed secondary structure of proteins it is important to understand the "weak" or secondary interactions (bonds) which determine the conformation assumed by a given polymer

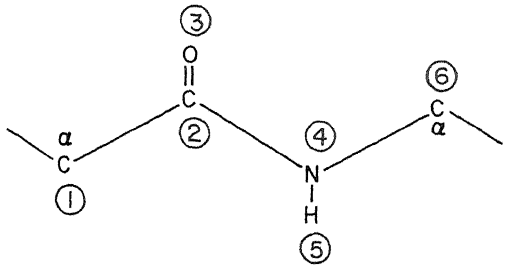


Fig 1 109 Characteristics of a peptide bond as revealed by x-ray diffraction

OTHER TYPES OF BONDS

Covalent bonds generally have energies between 50 to 110 kcal /mole Consequently they are stable at physiological temperature (37° C), they have relatively short bond lengths, and the number of bonds which can be formed between atoms is determined by the valence Furthermore bond formation proceeds under the aegis of enzymes In contrast, secondary (noncovalent) bonds have energies between 1 to 7 kcal /mole, therefore, a single secondary bond exists only for a very short period of time at 37° C Finally, secondary bonds can be formed and disrupted without the intervention of enzymes This latter characteristic is of no small consequence for cellular economy and is of central importance in the process of self-assembly of macromolecular complexes from subunits

Weak bonds are essentially interactions involving the electron clouds around atoms or groups of atoms, these interactions fall into three classes the van der Waal's bond, the hydrogen bond, and the ionic bond (A fourth class of weak interaction, the "hydrophobic bond," is discussed below, however the hydrophobic interaction is not a true bond in the chemical sense)

The van der Waal's bond is the weakest It has an energy of approximately one kcal /mole (random thermal motion in a solution at 37° C is about 0 6 kcal /mole) This bond results from the synchronization of the motion of the electrons around two approaching atoms or groups Each electron cloud induces a complimentary dipole in the other This is a very short range

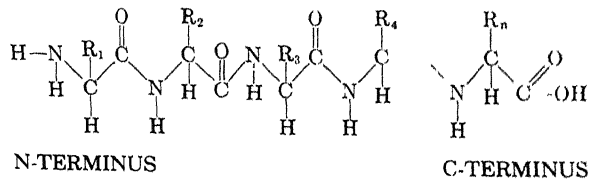


Fig 1 110 Generalized structure of a protein showing peptide bonds, N-terminus and C-terminus

force The energy falls off as the sixth power of the distance between interacting centers (i e , E ∞ $[1/r^6]$) As the interacting centers get closer, the attractive force is opposed by the repulsive force generated by the proximity of the electron clouds around the atoms This can be seen in a plot of energy vs distance (Fig 1 111) The distance r_v is the van der Waal's radius It has a characteristic value for each atom or group involved in an interaction This distance is a compromise between the proximiy necessary for the attractive force to be generated but not counteracted by the repulsive force resulting from the overall negative charges of the electron clouds

Inherent in the requirement for proximity is a steric requirement In order for the interacting groups to approach the critical distance needed for the induction of the attractive force, there must be a complimentarity of overall structure in the surfaces of the molecules to which the interacting groups or atoms are attached If this does not obtain, the bond will not form since the groups will be too far apart Lastly, since the forces are weak, the van der Waal's interaction will be more effective when several atoms or groups participate

A hydrogen bond is based on the attraction of unlike charges When a hydrogen atom is covalently bonded to a moderately strong electronegative atom (or group), it assumes a partial positive charge because the molecular orbital of the electrons is displaced in the direction of the electronegative atom, thus decreasing the neutralization of the positive charge at the site of the hydrogen nucleus This partial positive character allows the hydrogen to interact with any negative atom in another molecule to form a hydrogen bond (Fig 1 112) A-H is the donor species, B-C the acceptor species

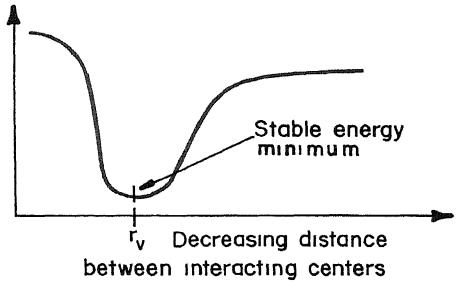


Fig 1 111 Plot of energy as a function of distance or bonds of van der Waal s type

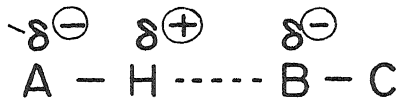
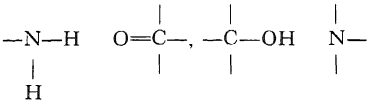


Fig 1 112 Diagram showing hydrogen bond

The following are some examples of common hydrogen bonds



Thus, any covalently bonded hydrogen with a partial positive character can interact with a covalently-bonded atom that has a net negative charge or unshared electrons The result of this interaction is the joining together of the two molecules via this hydrogen bridge This same interaction also occurs intramolecularly This results in two portions of a chainlike molecule being held in close proximity An example of the *intermolecular* hydrogen bond is that which holds two strands of DNA together An example of the *intramolecular* hydrogen bond is that which participates in the maintenance of the coiled superstructure of proteins (Fig 1 113)

The polar nature of water molecules leads to the formation of intermolecular hydrogen bonds and, accordingly, to a water structure often referred to as "iceberg structure " These intermolecular H bonds make and break rapidly, thus, the water structure is transient although its interaction with proteins is one of the factors that serve to determine the overall conformation of proteins

Hydrogen bonds have energies between 3 to 7 kcal /mole, thus, they are stronger than van der Waal's bonds Nonetheless, at physiological temperatures, they do not persist singly for a long time However, when present in ordered groups of molecules such as proteins or DNA, they have a longer lifetime and exhibit the characteristics of cooperative interactions This is a statistical phenomenon At physiological temperatures, weak bonds are in a state of flux, i e , they are constantly being formed and disrupted. However in the native state there are at any time a sufficient number of intact bonds that serve to maintain secondary structure If the temperature is raised more bonds are broken per unit time than are formed Further-

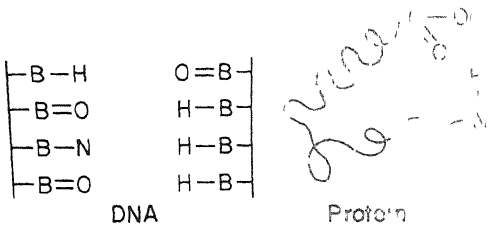


Fig 1 113 Hydrogen bonds in DNA and in a protein molecule

more, as the bonds break, it requires less energy to disrupt the remaining ones This causes an "unzipping" effect, which along with other factors, results in a "melting" of the polymer superstructure as shown in Figure 1 114 Figure 1 114 is a graphic representation of cooperativity In the temperature range T_1 to T_2 there is a rapid transition due to many bonds breaking per unit time for a small temperature increment Approaching T_1 from the left, it requires a larger temperature increment to start the process because initially there are many intact bonds Beyond T_2, most of the bonds are disrupted and further input of energy elicits a lesser response

Ionic bonds (salt bridges) occur between covalently bound atoms that are oppositely charged These bonds or bridges can be formed intramolecularly or intermolecularly, as shown in Figure 1 115, *a* and *b* respectively

Of the three classes of weak interactions discussed, the hydrogen bond is the one involved in the secondary structure of proteins

CONFORMATION OF PROTEIN MOLECULES

The most stable conformation of a polypeptide backbone requires that the carbonyl oxygens and amide nitrogens be hydrogen bonded, the peptide bond be planar, the hydrogen bonded unit be linear, and the geometrical operation in going from one residue to the next be the same along the whole chain There are two structures which meet these requirements, the α-helix and the β-pleated sheet (β-structure)

The backbone of a protein consists of a linear repetition of peptide groups In a regular linear polymer, the monomers are capable of forming sets of secondary bonds of varying strengths The energy differences between these sets of bonds will result in an unstable state for the overall polymer. However, if every monomer can be placed so that each is oriented identically within the molecule, then each monomer will form secondary bonds with identical energy contents Hence, helical symmetry is a result of the unique monomer arrangement which is the most stable of all of the possible arrangements

The α-helix (Fig 1 116) has the following characteristic parameters

1) Hydrogen bonds are parallel to the central axis of the helix
2) Hydrogen bonding occurs between the carbonyl group of the first amino acid residue and the amide group of the fourth residue between the carbonyl of the second residue and the amide of the fifth residue etc
3) The vertical distance corresponding to one turn of the helix i e the pitch of the helix is 5 4 Å
4) A complete spiral requires 3 6 residues
5) Therefore the vertical translation per residue equals 5 4/3 6 or 1 5 Å

The α-helix results from the formation of the maximum number of *intramolecular* hydrogen bonds The β-structure (pleated sheet) also satisfies the requirements for maximum conformational stability (Fig 1 117) However, in this case the hydrogen bonds occur *intermolecularly* between the carbonyl and amide groups on extended polypeptides that lie side by side Consequently the hydrogen bonds are perpendicular to the long axis of the backbone This gives rise to sheetlike structures which may involve parallel or antiparallel arrays of peptide strands

The foregoing considerations were based on the studies of crystalline amides and peptides However, secondary structure can also be demonstrated in polypeptides in solution Hydrogen bonding causes alterations of the dissociation constants of the involved groups, spectral changes and alteration in the rate at which

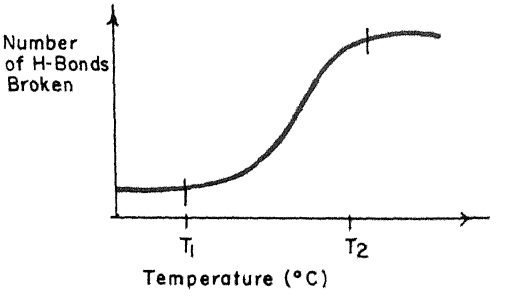


Fig 1 114 Stability of hydrogen bonds as a function of temperature

deuterium (D) will exchange for hydrogen in the protein, because an H atom involved in a hydrogen bond will exchange with D more slowly or not at all Measurement of the above parameters serves to indicate the H-bonding and the degree of secondary structure of a protein in solution

The interactions of the amino acid side chains (R groups) attached to the coiled or pleated peptide backbone—with each other and/or with the solvent environment— determine the next level of geometric complexity and give the overall molecule its final three-dimensional character (*tertiary structure*) Thus, the tertiary structure defines the steric relationship of each atom to every other atom in the molecule

There is a fourth class of weak interactions which is of major importance in determining the final three-demensional conformation of the protein molecule, the hydrophobic "bond " This is not a true bond in the chemical sense Rather, it represents the tendency of hydrophobic (nonpolar) residues to associate together, away from the aqueous (polar) environment This gives rise to apolar clusters, i e , centers in the polymer where water is excluded

This behavior can be understood by considering the components of an aqueous protein solution Such a solution is composed of polar

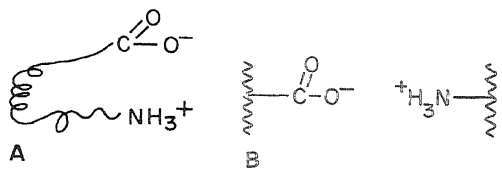


Fig 1 115 Intramolecular (*a*) or intermolecular (*b*) ionic bonding

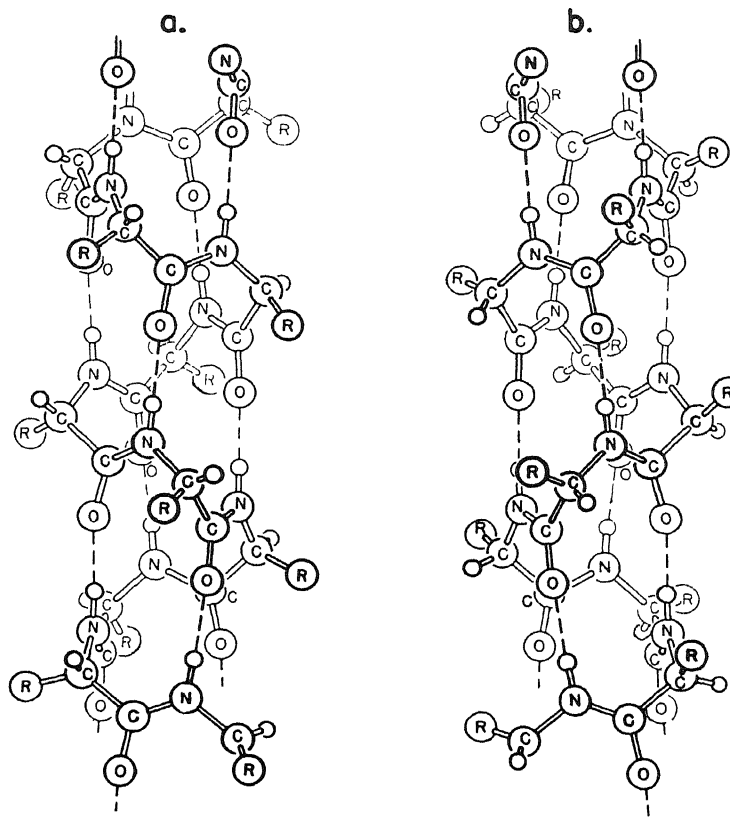


Fig 1 116 Drawings of the left-handed (*a*) and right-handed (*b*) α-helical forms of a polypeptide chain containing L-amino acids The side chains *R* and hydrogen atoms attached to the α-carbons in the main chain have positions corresponding to the known configuration of L-amino acids (From Low and Edsall 1956)

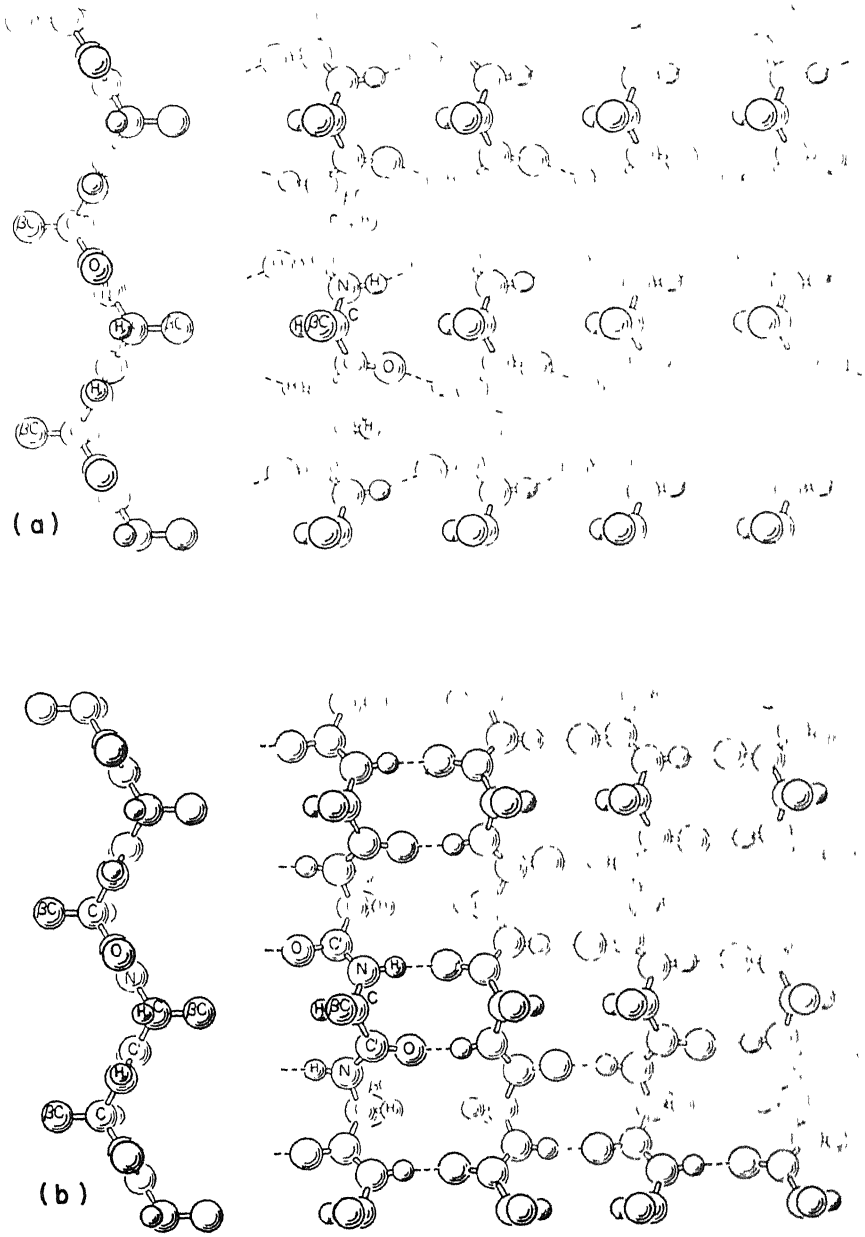


Fig 1 117 Representation of the parallel (*a*) and antiparallel (*b*) pleated sheet structures for polypeptides (From Pauling and Corey 1951)

solvent molecules (water), polar side chains (R groups) of amino acids, and apolar amino acid R groups The polar R groups can interact and form H bonds with the iceberg lattice of water Each R group becomes surrounded by a cage of water molecules but this ordering effect on the aqueous solutions decreases with the distance of the water molecules from the R group The lattice structures thus formed are transient, and the individual water molecules are free to move within the lattice On the other hand, nonpolar groups do not interact or form H bonds with the water lattice but rather serve to "lock" the water lattice structures and impede the mobility of free water molecules This constitutes a decrease in entropy which requires an

expenditure of free energy, a circumstance which is in contrast to the general tendency of systems to minimize their energy content and maximize their entropy ($\Delta G = \Delta H - T\Delta S$)

The amino acid side chains can be considered as polar and nonpolar A polymer composed of a fixed ratio of polar to nonpolar side chains in aqueous solution can theoretically exhibit several conformations each of which has a different energy content, as illustrated in Figure 1 118 The conformation with the lowest overall energy is the most stable Consequently, A (Fig 1 118) will be the longest lived, and hence, the predominant conformation However, let us introduce another polymer which contains the same total number but a *slightly* larger proportion of nonpolar side chains Now, if only the backbone is considered, the most stable conformation will be one in which its secondary bonds are maximized, i e , a helix However, if our consideration is extended to include the side chains, the most stable structure will no longer be a helix but rather a new conformation resulting from a compromise between the tendency to form a helix being offset by the tendency of the side chains to twist the backbone so that nonpolar groups can associate The compromise conformation usually favors the maximization of secondary bonds for the side chains (approximately 50% of the amino acid side chains are nonpolar)

Since peptide sequence is random, nonpolar side chains can occur in widely separated locations along the backbone However, when these side chains come into proximity, they form hydrophobic bonds These nonpolar groups usually place themselves in the interior of the protein molecule, leaving the polar side chains outside to interact with the solvent To bring these groups out of their hydrophobic pocket, where water is excluded, and to force them into a polar solvent (water) requires a large increase in free energy expenditure because of the decrease in entropy of the system

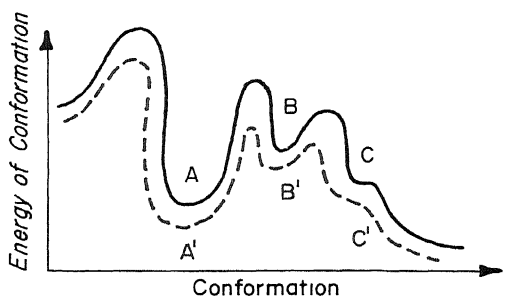


Fig 1 118 Variability of energy of conformation (ordinate) of different conformations of a polymer in aqueous solution

Thus, conformation A' (Fig 1 118) is more stable than conformation A, more energy being required to disrupt the structure and force the nonpolar side chains into the aqueous environment in which they cannot interact with each other or with the water This would produce a large decrease in entropy (hence, the necessity for a large energy expenditure) However, conformation A contains fewer apolar residues, consequently the magnitude of the decrease in entropy associated with disruption of A would be less than that associated with disruption of A'

Therefore, the tertiary structure of a protein is the thermodynamically most stable arrangement It represents the compromise result of simultaneous maximization of polar, hydrophobic, and hydrogen-bonding interactions

The hydrogen bonding of the peptide backbone involves interactions between one part of the protein molecule and another (intramolecular interactions) and between the protein molecule and the solvent (intermolecular interactions) When because of the hydrophobic interaction of the side chains, the protein molecule is so twisted that the peptide linkages are forced into the interior of the molecule with little access to solvent, the intramolecular reactions will be maximized and the α helical conformation is generated Conversely those regions of peptide backbone which are not interiorized, and therefore can form hydrogen bonds with the solvent will assume the conformation of a random coil

The thermodynamically most stable state of a protein, its tertiary structure, is its native conformation This is the conformation that exhibits biological activity This concept has been verified in the case of the enzyme, ribonuclease

Loss of tertiary structure in a protein is brought about by any factor or combination of factors which disrupt its weak interactions (secondary bonds) and is termed *denaturation* Denaturation can be accomplished by manipulations which change a highly ordered structure into a random one, such as heating, drastic pH changes, shaking, organic solvents, detergents and irradiation A denaturated protein lacks

biological activity, denaturation can be reversible (as is the case following treatment with urea) or irreversible (as is the case following heat treatment of proteins) The denaturation process can be followed in solutions of proteins by observing spectral changes (optical rotation ultraviolet absorption at 280 mμ) and changes in solubility (precipitation)

When the enzyme ribonuclease is denatured by reductive cleavage of its disulfide[1] bridges and by treatment with urea, it loses its biological activity However, if the urea is removed from the denatured preparation and then its cysteine moieties are reoxidized to reform disulfide bridges, biological activity is restored Furthermore, analysis of the resulting structure indicates that the disulfide bridges assume their original positions Thus, prior to reformation of the disulfide bridges, the random (denatured) structure must convert to the native conformation upon removal of the urea The basis for this conclusion rests upon the fact that there are four disulfide bridges (eight half-cystine residues) in the native conformation at specific loci in the molecule If the native state were not obtained prior to reoxidation of the eight cysteine residues the disulfide bridges would be reformed at random In this experiment there was no agent present in the solution to direct the conversion of the denatured enzyme to its native conformation The basis for this renaturation would appear to be solely the characteristic property of proteins to assume the thermodynamically most stable state Thus, the information present in the primary structure of ribonuclease will dictate the most stable conformation of the molecule, i e , the sequence of the amino acids in the molecule will determine the type, number, and location of the secondary bonds This is an extremely important consideration inasmuch as it explains how primary amino sequence in conjunction with a very limited number of secondary interactions—viz , hydrophobic bonding, hydrogen bonding, ionic bonding—serves to establish the many protein conformations associated with living systems All proteins existing in the same environment, with the same amino acid sequence must have the same conformation Changing the amino acid sequence establishes a different conformation by establishing a different state of maximal thermodynamic stability This new conformation may be associated with complete or partial loss of biological activity or rarely, with the same or enhanced biological activity The ramifications of these considerations will be further explored in the discussion of the genetic code and protein synthesis

In summary, the peptide backbone of a protein is sufficiently flexible to allow the molecule to assume a large number of different conformations (conformers) However secondary interactions, particularly involving the amino acid side chains restricts the ultimately realized conformation to one that is unique for the primary sequence of the protein and its environment These secondary interactions may lead to different ultimate conformations for different protein molecules or for different regions within a single protein molecule, e g , helices, random coils or compact configurations established by bending and folding of the peptide backbone Such bending and folding, which characterize the globular proteins, bring into contact amino acid residues that may be widely separated in the primary structure[2]

In addition, nonpolar groups located in the interior of protein molecules give rise to hydrophobic pockets which can be important determinants of the molecular complementary that underlies specificity in hormone-receptor, enzyme-substrate and antigen-antibody interactions and also can play a role in the electronic aspects of enzymatic reactions, e g., by amplifying the effects of charged groups in their vicinity (by excluding the dielectric effect of water)

QUATERNARY STRUCTURE

Proteins (such as hemoglobin, the immunoglobulins, the allosteric enzymes, etc.), consisting of more than one chain per molecule are said to have a *quaternary structure* The forces involved in holding the multiple polypeptide chains (subunits) together in a single molecule are the same as those which establish the sec-

[1] Disulfide bridges are low energy covalent sulfur-sulfur bonds formed between 2 cysteine residues, i e ,

$$2\ \text{CYS-SH} \underset{\text{reduce}}{\overset{\text{oxidize}}{\rightleftharpoons}} \text{CYS-S-S-CYS}$$

The disulfide bridge is an important component of the tertiary structure of many proteins, adding strength and support to the three-dimensional architecture (conformation) of the molecules

[2] This phenomenon may be of importance in the combination of antigen and antibody molecules where antigen-induced "additional" folding of the antibody polypeptide chains may be required for the formation of the antigen-antibody complex

ondary and tertiary structure They operate under the same set of thermodynamic restrictions Thus, each subunit of the functional protein possesses its own conformation and is capable of interacting with another subunit The subunits can be identical or different What is required for the secondary interactions to operate is a complimentarity of surface configurations so that reactive sites on subunits can approach each other closely enough for the formation of ionic bonds, hydrogen bonds, hydrophobic interactions, and interchain S-S bonds The overall active conformation of the protein aggregate is specified by the conformations of the individual subunits Once the subunits exist in their thermodynamically most stable configuration, they then associate *spontaneously*, giving rise to the active aggregate protein This process is referred to as self-assembly It is not mediated by enzymes and does not require an external energy supply It can account for the elaboration of giant macromolecular complexes For example, lactic acid dehydrogenase and aspartyl transcarbamylase are such macromolecular enzyme complexes Virus coat proteins also follow this mode of assembly, i e , once the subunits are synthesized, they spontaneously aggregate around the genetic material

From the viewpoint of cellular economy this is an important consideration It means that no enzymes are required to direct assembly Furthermore, if a subunit is defective it will not interact properly and, accordingly, it will be excluded from the superstructure On the other hand, if a defect develops in a macromolecular superstructure which is synthesized as a single polypeptide chain, a complete breakdown and resynthesis would be necessary for repair of the defect This process can be viewed as a molecular assembly line The subunits are synthesized and then joined to yield the finished functional product

Since the subunits are held together by weak interactions they can be dissociated reversibly by mild procedures such as changing pH or ionic strength In the dissociated state the protein is biologically inactive, aggregation leads to a recovery of biological activity

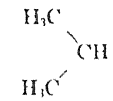
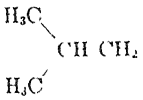
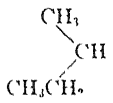
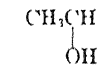
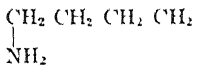
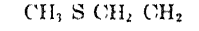
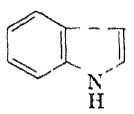
The subunit construction of a giant enzyme complex facilitates the allosteric regulation of its function, e g , a given subunit may form a complex with a small molecule, the allosteric effector or repressor, thereby inducing a conformational change in the entire aggregate which leads to activation or inhibition, respectively, of enzymatic activity

SOURCES OF AMINO ACIDS IN THE METABOLIC POOL

Protein ingestion provides an *exogenous* source of amino acids Before discussing the metabolism of amino acids it is important to enumerate some of their nutritional parameters Protein nutrition must be considered from the standpoint of the amino acids which constitute the ingested protein There is a qualitative as well as quantitative requirement In a classical study Osborn and Mendel kept groups of rats on various diets of purified proteins and measured the ability of these proteins to support growth (Mendel, 1923) Rats maintained on gliadin or zein (wheat and corn protein, respectively) exhibited poor growth, whereas, rats maintained on ovovitellin (from egg yolk) or lactalbumin (from milk) exhibited excellent growth The latter groups of animals grew well because the ingested protein contained all *essential amino acids*, viz , amino acids which are required for vital anabolic reactions, but which cannot be synthesized endogenously from readily available precursors at a rate sufficient to meet the metabolic needs of the animals Essential amino acids (Table 1 9) must be included in the diet or they will be commandeered from the animals' own tissue protein Thus, an inadequate intake of essential amino acids results in deficient protein synthesis, net tissue catabolism and ultimately death of the animal—regardless of the absolute quantity of protein ingested Diets containing proteins which are deficient in one or more essential amino acids can be supplemented with either the deficient amino acids or, more feasibly, with other proteins that do not lack these amino acids This is the reason that the diet should contain a wide range of foods and that those people who choose to or are forced to restrict their diets to food of a single type (such as rice or corn) are likely to develop protein deficiency syndromes

Another determinant of the qualitative and quantitative protein requirements is the physiological state of the organism These requirements vary with age, sex, disease, nutritional state, etc Thus, histidine is an essential amino

Table 1 9
Amino acids essential for normal human nutrition

Amino Acid	R Group (Side Chain)
Valine (Val)	
Leucine (Leu)	
Isoleucine (Ile)	
Threonine (Thr)	
Lysine (Lys)	
Methionine (Met)	
Phenylalanine (Phe)	
Tryptophane (Try)	

[a] Histidine (His) is essential only for infants and young children Tyrosine (Tyr) can be synthesized only from phenylalanine (Phe), therefore a deficiency of Phe renders Tyr essential and, conversely, an ample supply of Tyr reduces the requirement for Phe because synthesis of Tyr is not necessary The relationship between cystine (Cys) and methionine (Met) is similar, for cystine can be synthesized from methionine which, as noted above, is essential

acid for infants and small children but not for adults (the reason for this difference is unknown) Menstruation, lactation, and pregnancy are associated with an increased quantitative requirement for all amino acids, essential and nonessential

For amino acid utilization to be optimal, the diet must include a sufficient caloric intake to circumvent the need to catabolize amino acids for energy, thus, allowing these compounds to be used in synthesis Fat and carbohydrate fulfill this function and, therefore, are referred to as *protein sparers* Thus, it can be seen that particular amino acids along with an adequate caloric intake are essential for normal protein metabolism Over and above this, the protein must be made available for utilization by the tissues i e there must be adequate hydrolysis by the proteolytic enzymes of the digestive tract because absorption of intact protein does not occur under normal conditions In some situations small quantities of protein do gain access to the circulation a circumstance which can lead to sensitization and in turn to the development of hypersensitivity (allergy) to the particular food from which the sensitizing protein derives (e g strawberries) In other situations, blood proteins are lost by abnormal leakage through the intestinal epithelium

After ingested protein is hydrolyzed the amino acids must then be absorbed efficiently into the portal circulation The separate functions of digestion and absorption become manifest in certain pathological states For example in a condition characterized by deficiency of the pancreatic proenzyme, trypsinogen, malnutrition ensues because of failure of digestion (maldigestive malnutrition) This condition can be remediated either by trypsinogen replacement, which corrects the digestive failure, or through supplementation of the diet with protein hydrolysates, thus bypassing the initial enzymatic deficiency

The absorptive cells of the small intestine accumulate free amino acids from the gut lumen (concentrative uptake) and then deliver them to the portal circulation These transport mechanisms are active and possess specific physical and chemical characteristics which are under genetic control In vivo as well as in vitro studies indicate that a single amino acid utilizes more than one mode of transport for absorption

The cooperation of several transport systems is required for the absorption of an amino acid mixture and it is through these several systems that specificity is obtained Each system can handle several amino acids; consequently the various individual amino acids handled by a given system will compete with each other for transport from the gut lumen to the plasma Thus, the absorption of an amino acid can be affected by back transport of a related amino acid from the plasma to the gut lumen. Further characteristics of transport systems will be discussed subsequently when pathological conditions are considered

The *endogenous* input of amino acids into the plasma derives from the breakdown of tissue proteins (effected by proteolytic enzymes called cathepsins) and from the synthesis of nonessential amino acids primarily by liver

cells Free amino acids from all sources enter the plasma, however the plasma concentration remains relatively constant This constancy of plasma amino acid concentration is due to the regulation of and interrelationships between the complex transport and metabolic systems For example, amino acids introduced into the portal circulation as a consequence of protein ingestion can increase the concentration of amino acids in the portal vein by 10-fold This increase is not reflected in the systemic circulation because there is an active uptake of amino acids by the tissues leading to an intracellular free amino acid concentration 10 times greater than the plasma concentration

Amino acid transport processes regulate absorption, metabolic equilibrium, and renal conservation The transport mechanisms, in turn, are regulated by hormones and nuclear cytoplasmic interactions Since the intracellular concentration of amino acids is greater than the extracellular concentration, cellular metabolism has access to a greater pool of metabolites than is available from the immediate extracellular source Metabolic changes in the internal pool can trigger compensatory changes by modifying the transport systems In this way, amino acids are either taken up or discharged from the cell The interrelations among the various pools is summarized in Figure 1 119

INTERMEDIARY METABOLISM OF AMINO ACIDS

Amino acid metabolism is a diverse repertoire of reactions which form a loom upon which the structure and enzymatic machinery that govern the chemical reactions of living systems are woven into their characteristic patterns Diversity stems from the variety of the carbon skeletons of the various amino acid side chains and the fact that amino acids supply the major source of fixed nitrogen for the body To handle this class of metabolites requires a large number of enzymes and an elaborate system of controls The various carbon skeletons are metabolized by sets of their own specific enzymes Handling of the nitrogen moiety requires other enzymatic apparatus which determines whether the nitrogen is transferred to amino acid precursors, fixed in nonprotein nitrogenous compounds, fixed in proteins, or eliminated from the body

All amino acids contain C, H, O, and N, some contain S The catabolism of amino acids results in CO_2, H_2O, nitrogenous end products (principally urea) and sulfates and, in addition, is coupled to the generation of adenosine triphosphate (ATP) However, the primary importance of the amino acids is not as metabolic fuel, but in the synthesis of proteins and of

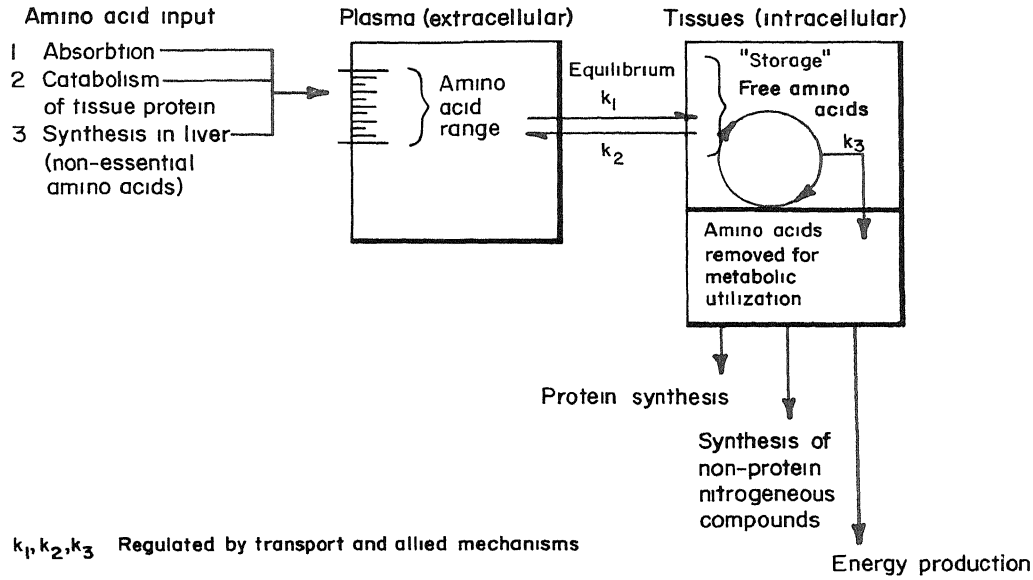


Fig 1 119 Interrelations among the several pools of amino acids within the body

other nitrogen containing compounds. (See Fig. 1.120 and Table 1.10).

The catabolism of amino acids can be considered in terms of the fates of the carbon and nitrogen moieties. The first stage of catabolism involves the removal of the amino group (deamination). The resulting carbon skeleton is then metabolized as an intermediate of either the Krebs cycle or of fatty acid oxidation cycles. Amino acids that yield intermediates of the former cycle are called glycogenic, those that yield intermediates of the latter cycle are referred to as ketogenic (see Table 1.11). Glucogenic amino acids provide carbon skeletons for the synthesis of glucose (gluconeogenesis) in the liver and kidney. This process is under hormonal control (glucocorticoids) and helps energy production keep pace with utilization.

Another aspect of amino acid catabolism is the poorly understood phenomenon termed the *specific dynamic action* or *calorigenic effect* of food. This refers to an extra increment of heat production over and above the caloric content of the food ingested. Fats, carbohydrates, and proteins all evoke this action. However, proteins produce a larger increment, viz., up to 30% more heat can be generated than can be accounted for by the quantity of protein ingested. Fats and carbohydrates by comparison yield an increment of only a few per cent.

Decarboxylation of amino acids yields the corresponding amines. Some of these compounds are important physiological entities with hormonal, neurotransmitter, and other functions (e.g., epinephrine, serotonin, etc.; see Table 1.8); others are toxins (e.g., the action of bacteria on incompletely digested *amino acids* yields toxic amines referred to as ptomaines). Decarboxylation reactions are mediated by specific amino acid decarboxylases. These reactions require pyridoxal phosphate (codecarboxylase) as a cofactor. The general reaction is represented in Figure 1.121.

The nitrogen moiety of amino acids has essentially two fates in the body. It can be removed from the amino acid (deamination) and eliminated from the body, principally as urea. Alternatively, it may be transferred from an amino acid to an α-keto acid, thus forming a different amino acid and leaving behind a different α-keto acid. These reactions, which are referred to as *transamination*, are carried out primarily in the liver.

There are two systems which are involved in the oxidative deamination of amino acids. The flavin adenine dinucleotide-(FAD) dependent L-amino acid oxidase system has a low activity and is not widely distributed in the body. Also, it does not metabolize dicarboxylic, diamino or

Table 1.10
Some physiologically significant nitrogen-containing compounds and the amino acids from which these compounds are derived

Amino Acid(s)	Nitrogenous Compound
Tyr	Epinephrine, norepinephrine, thyroxin, melanin, dihydroxyphenylalanine
Glu	Pyrroles, porphyrins, heme
Gly, Glu, Asp	Purines, pyrimidines
Gly, Arg, Met	Creatine, creatinine
Cys	Ethanolamine
Try	Taurine, nicotinic acid, 5-OH-tryptamine, serotonin

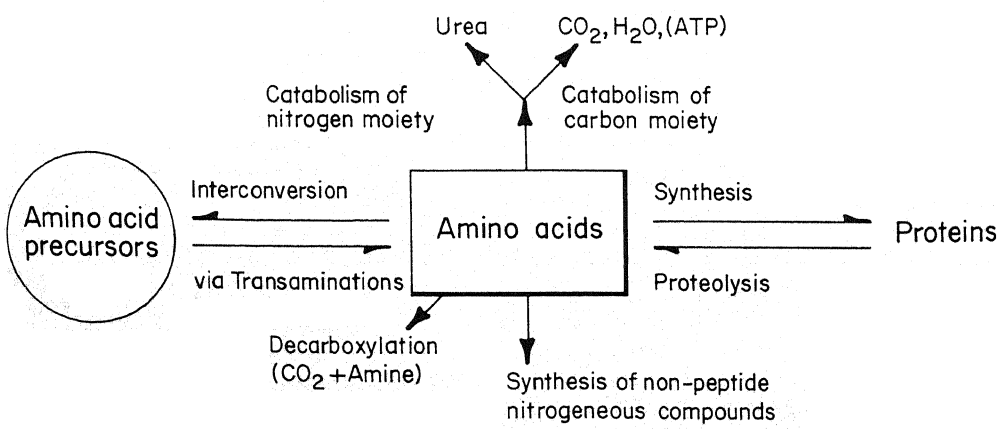


Fig. 1.120. Metabolic pool of amino acids as a source of proteins and other nitrogen-containing compounds.

β-hydroxy amino acids and glycine The enzymes of this system catalyze the general type of reaction shown in Figure 1 122 D-amino acid oxidases are also present in liver and kidney, however, their physiological role is obscure

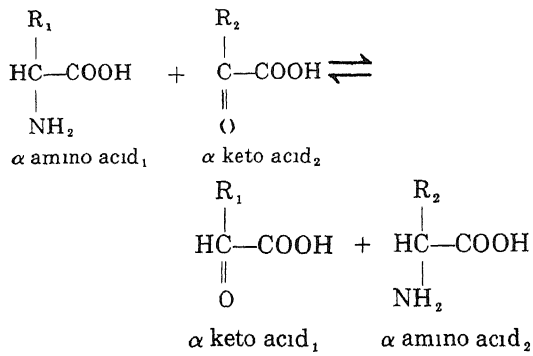
The second, more important system for oxidative deamination has a wide distribution and high activity This is the L-glutamic acid dehydrogenase system It is the major pathway for deamination and is central to amino acid interconversions This is an nicotinamide adenine dinucleotide-(NAD) dependent system and catalyzes the general reaction shown in Figure 1 123 The reversibility of this system is its essential feature as will become apparent in the discussion of transamination reactions (see below)

The amino acid, glycine, is deaminated to glyoxylic acid and NH_3 by a special system involving glycine oxidase

Special systems also exist for the deamination of the hydroxy-amino acids, serine, threonine, and homoserine These are nonoxidative deaminations The enzymes catalyze primary dehydrations which lead to spontaneous deaminations, pyridoxal phosphate is required as a cofactor

A second type of nonoxidative deamination involves the sulfur containing amino acids cysteine and homocysteine The primary reaction is loss of H_2S (desulfhydration) resulting in an amino acid which spontaneously hydrolyzes to pyruvic acid and ammonia Again this system requires pyridoxyl phosphate as a cofactor

All amino acids with the exception of threonine and lysine, have been shown to undergo transamination reactions Since these are intermolecular transfers of the nitrogen moiety, NH_3 is not generated Transaminase systems require pyridoxal phosphate as a cofactor The general reaction can be summarized as follows



Glutamic acid is a common intermediate in the various transamination systems The forward

Table 1 11
Glycogenic and ketogenic amino acids

Glycogenic	Ketogenic	Either Glycogenic or Ketogenic
Alanine	Leucine	Isoleucine
Arginine		Lysine
Aspartic acid		Norleucine
Cysteine		Phenylalanine
Glutamic acid		Tyrosine
Glycine		
Histidine		
Hydroxyproline		
Methionine		
Ornithine		
Proline		
Serine		
Threonine		
Tryptophan		
Valine		

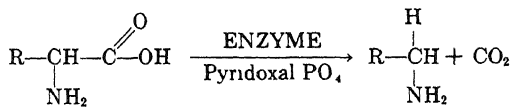


Fig 1 121 Decarboxylation of amino acid to yield an amine

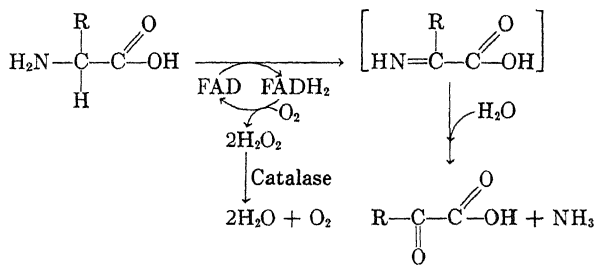


Fig 1 122 Reactions of the L-amino acid oxidase system

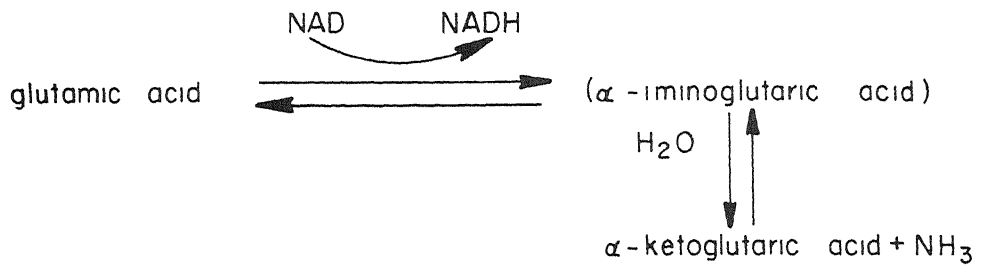


Fig 1 123 Reactions of L glutamic acid dehydrogenase system for deamination of amino acids

reaction removes amino groups from other amino acids (R_1) forming glutamic acid (R_2) via transamination Glutamic acid (R_2) can then act as a NH_2 donor to transaminate keto acid precursors to form other amino acids Thus, the reversible deamination of glutamic acid to α ketoglutaric acid and NH_3 is a central reaction for amino acid synthesis as well as catabolism In addition this reaction links carbohydrate and protein metabolism through the Krebs cycle (α-ketoglutarate)

Glutamate, aspartate, or alanine seems to be common to every reaction This means that the NH_2 group from any other amino acid can appear in any of these three amino acids NH_2 entering glutamate can be removed via the glutamic dehydrogenase system This provides a pathway for deamination of a wide range of amino acids This "transdeamination" pathway results in formation of α-ketoglutarate and NH_3

The following reactions represent the above transamination systems

alanine + α-ketoglutarate $\longleftrightarrow$ pyruvic acid + glutamic acid

glutamic acid + oxaloacetic acid $\longleftrightarrow$ α-ketoglutarate + aspartic acid

alanine + oxaloacetic acid $\longleftrightarrow$ pyruvic acid + aspartic acid

In certain diseases (myocardial infarction, liver diseases), damaged cells release transaminase systems into the blood stream Consequently, the levels of activity in the serum of glutamic-oxaloacetic transaminase (SGOT) and other transaminase systems are increased This increase is proportional to the extent of damage and, therefore, assays of the level of activity of SGOT and/or other transaminases are of clinical significance

Free NH_3 results from deamination of amino acids, pyrimidines and purines, from the deamidation of glutamine, asparagine (and their α keto analogues) and to some extent from the action of urease-containing intestinal bacteria on urea (urea $\xrightarrow{\text{urease}}$ CO_2 + NH_3) Free NH_3 in high concentrations is incompatible with life Ammonia must therefore be either bound in the organism or excreted It has already been shown that NH_3 may be bound by the "amination" of α-keto acids Amidation of glutamate and aspartate to form glutamine and asparagine, respectively, are other reactions that bind free NH_3 for transport in the plasma ATP is required for the glutamine synthetase reaction In the kidney, glutamine is converted by glutaminase to glutamate and ammonia, the latter diffusing into the tubular urine where it combines with H^+ to form NH_4 The renal tubular cells also contain glutamate dehydrogenase and transaminase systems as well as amino acid oxidases, therefore, the kidney can also generate NH_3 from other amino acids

Some ammonia is excreted as uric acid, however, the major portion of NH_3 is excreted by conversion to urea The enzymes of this detoxication system are present in the liver and constitute the Krebs-Henseleit cycle (Fig 1 124)

Amino nitrogen can enter this cycle either through fixation with CO_2 to form carbamyl phosphate (I), or glutamate can transaminate oxaloacetic acid to yield aspartate (II) Each revolution of this cycle converts 2 molecules of NH_3 and 1 molecule of CO_2 to urea

Amino acid metabolism is summarized in Figure 1 125

Overall protein metabolism is reflected by the nitrogen balance of the organism The nitrogen content of the ingested food is measured and compared with the nitrogen excreted in the feces and urine over a given period of time *Positive nitrogen balance* is obtained when the total nitrogen intake is greater than the total nitrogen output, as is the case when anabolic reactions (principally protein synthesis) predomi-

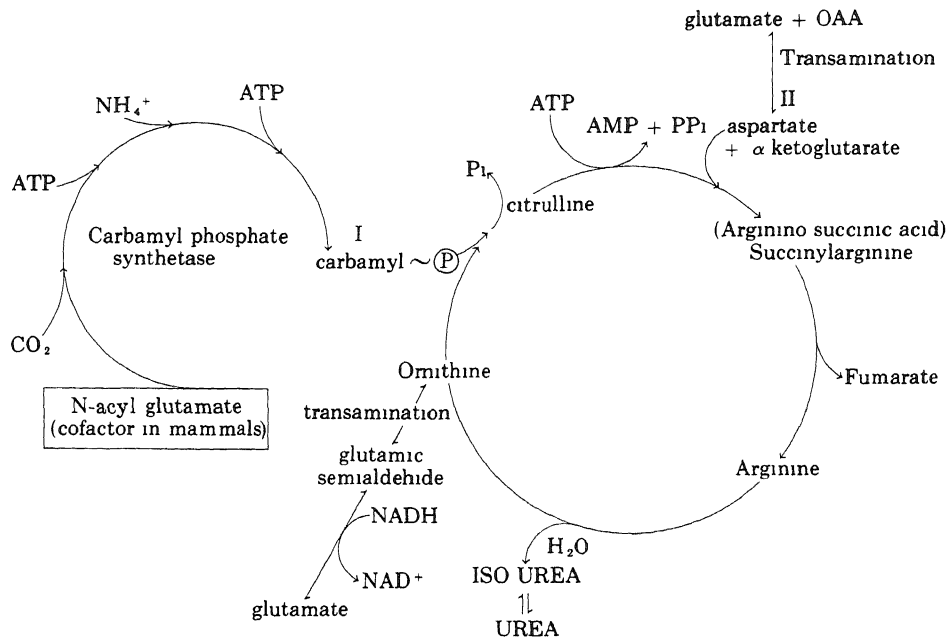


Fig 1 124 Krebs-Henseleit cycle

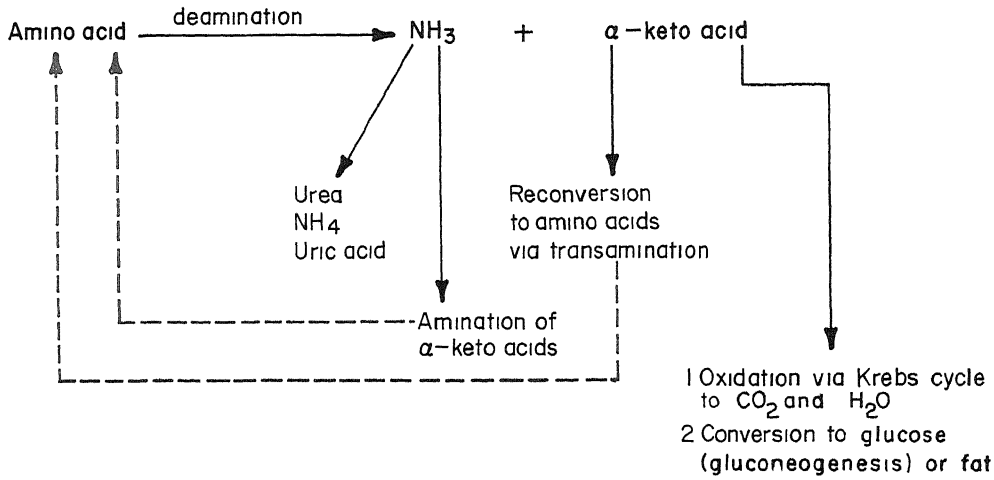


Fig 1 125 Summary of chemical reactions through which amino acids are metabolized

nate over catabolic reactions, e g , during growth or during periods of tissue repair after trauma, surgery, or illness If catabolic reactions exceed anabolic reactions, nitrogen excretion exceeds intake and the organism is in *negative nitrogen balance* This occurs in wasting diseases and starvation, when tissues are being broken down to meet metabolic demands, and when the diet is deficient in essential amino acids If input equals output, catabolism equals anabolism and the organism is in *nitrogen balance* This is the condition for the normal adult

Aminoacidopathies and Proteinopathies

To date, there are over 50 disorders of amino acid metabolism which have been described It would not be appropriate here to detail these disorders with regard to specific biochemical lesion, metabolic consequence, clinical manifestation, and treatment Rather, a generalized

discussion of aminoacidopathies from the standpoint of genetics and metabolism will be developed

GENETIC BASIS OF MOLECULAR DISEASE

Inborn errors of metabolism or the more recent appellation—molecular diseases—are genetically determined defects in protein structure or function Thus in considering these diseases attention must be focused on the genotype as well as on its *expression*, i e , the phenotype On the molecular level, there is little difficulty conceptualizing a correspondence between DNA and polypeptides The cistron and its polypeptide are colinear and commensurate (3 bases per 1 amino acid) One cistron determines one polypeptide

The polypeptide can be a functional unit in itself such as the hormone adrenocorticotropic hormone (ACTH), or it may be a subunit which must combine with one or more other subunits to be active as in the case of hemoglobin However, cistrons which code for the subunits which comprise a functional biological entity (e g , enzyme complex, hormone, antibody, etc) need not be spatially contiguous Thus, the "gene" for such a functional biological entity can be composed of cistrons which are separated from each other along the length of the chromosome

The normally functioning organism is a result of the coordinated sequential expression of the information contained in its DNA Simultaneous expression of all genes would lead to cellular "anarchy " To circumvent this, there exists within the cell elaborate control mechanisms of nuclear-cytoplasmic interactions The resulting limitations on the number of gene products (polypeptides) found within any given cell results in specialization The organization of similar cells (i.e , cells with similar repertoires of reactions) into functional clusters yields tissues For orderly reactions to occur between tissues information must be exchanged among them This can be realized either by direct entry of one tissue product into the cells comprising a target tissue (e g , steroid hormones), or the tissue product may interact at the cell membranes of the target tissues resulting in the generation of a train of intracellular events, as exemplified by the action of a number of peptide hormones.

Interactions between systems are complex because all the systems are performing their own specialized functions simultaneously Thus, the history of a system constitutes one group of variables, among many, which determines the character of its response to a given stimulus (signal) Signals, therefore, may be amplified or ignored depending on the state of the system In addition, signals may be synergistic or antagonistic in evoking a response As the number of component systems increase, the complexity of the possible reactions increase This is ultimately reflected in the number and range of which the organism is capable It is the elaboration of many interacting control systems which integrate the activities of the organism to yield a response which is appropriate and graded to the intensity of the stimulus Indeed, it is the sum of these molecular control systems which characterize the organism biochemically and physiologically In viewing the organism macroscopically the molecular basis for its behavior is not always obvious since the stimulus-response relationship has been altered through the interaction of these multiple control systems

These control systems, in turn, are expressions of the genotype of the organism, which in its turn must be controlled Spatial and temporal regulation of the genotype is responsible for the character of the aforementioned systems and their behavior The expression of the full genotype in any one cell or tissue is never realized Only that portion of the DNA is expressed which results in products and activities that characterize particular cells or tissues

It will be recalled that the visible expression of the information in DNA is the phenotype, and also that particular traits, such as blood group, eye color, etc , are controlled by genes which exist in pairs (*alleles*), located on the same loci of the homologous chromosomes The alleles may be identical (*homozygous*) or they may be different (*heterozygous*) Usually, if they are heterozygous, one allele is dominant in the phenotype, the other is recessive (i e , the recessive allele is present in the genotype but not usually expressed in the phenotype) A well-known example is the dominance of brown eyes over blue eyes, which dictates that a child receiving a gene for brown eyes from one parent and its "blue" allele from the other parent will have brown eyes Thus, the child is heterozygous for eye color The expression of the recessive phenotype requires that the child be homozygous, i e , that he receive the reces-

sive "blue" allele from each of his parents Where there are more than two alleles capable of occupying the two homologous chromosomal loci that determine a given trait, this phenomenon is called *multiple allelism* Thus, eye color has many expressions, e g , brown, black, blue, green, hazel, etc , and is an example of the "continuous" variation that can result from multiple allelism and/or from partial expression of each of two different alleles This latter phenomenon has been termed *incomplete dominance* or *penetrance*

The foregoing treatment of the mode of expression of the phenotype of an organism is a simplification, but it provides a useful framework for the consideration of the transmission of traits through generations The "gene" is actually comprised of many cistrons which specify many enzymes which, in turn, govern the production of the ensemble of pigments that determine eye color These enzymes may be single polypeptides, subunit assemblies, or more probably mixtures of both—the full array of cistrons responsible for the trait constitutes the single inheritable entity which we recognize as the "eye color gene "

Another example of continuous phenotypic variation is height This trait, however, is the result of the interactions of many genes at different loci plus environmental influences such as nutrition Thus, the potential for greater height, although included in the genotype, may not be phenotypically expressed if environmental conditions militate against it, e g , inadequate dietary input

When differences in phenotype are not continuous but exhibit a quantal nature (i e , when there are two or more distinct forms of expression with no intermediate forms) then the phenomenon is referred to as polymorphism and exemplified by the ABO blood groups This is a case of multiple allelism in which there are three alleles, A, B, or O Various allelic combinations result in the four phenotypes—A, B, AB, or O—each of which is distinct from the other

The expression of any particular gene depends in part on the nature of its neighboring genes, thus, position on the chromosome per se can alter the functioning of a particular gene Events such as crossing over in meiosis can influence the immediate environment of a gene, and thereby alter or eliminate the gene's phenotypic expression, even though it remains in the genotype and its structure is unaltered For example, a gene located on a segment of a chromosome in the euchromatin condition of coiling, (i e , "normally" coiled and therefore translatable into a polypeptide) can after synapsis end up in a region of tightly coiled, nontranslatable heterochromatin, and thus have its expression in the phenotype suppressed Furthermore, the same gene may be expressed differently in different tissues, e g , the "transport system" for dibasic amino acids is found in the gut, kidney, and leukocytes but it functions differently and independently in each tissue—in the disease known as cystinuria the leukocyte system is normal, but the gut and kidney systems are abnormal

In humans the sex chromosomes have different morphologies—the X chromosome is larger than the Y chromosome XX is the female genotype, XY is the male genotype Some loci on the X chromosome are not represented on the Y chromosome and vice versa However, since the sex chromosomes are homologous they have some loci in common Since the female has two X chromosomes and the male only one, it would be expected that an enzyme determined by a gene located on the X chromosome would be present in twice the amount in the female as compared with the male But this is not the case For example, glucose-6-phosphate dehydrogenase is an enzyme determined by a gene located on the X chromosome It is conveniently assayed in red blood cells where it has been found to exhibit the same level of activity in both males and females This finding may be explained by a suppression of the action of one of the alleles in the female through coiling and clumping of the chromatin This concept, termed *dosage compensation*, is supported by cytological evidence in that diploid autosomal cells, neutrophils, exhibit a clump of chromatin known as the *Barr body* Since it is not found in the cells of the male, this is thought to be repressed sex chromatin

It is possible for more than one genotype to give rise to a single phenotype (*genetic heterogeneity*) Hemoglobins with structural differences can be grouped into classes with common characteristics, e g , altered electrophoretic mobility, altered ultraviolet spectrum and thermal instability The hemoglobins which comprise any one of the above "phenotypic" classes may have different amino acids substituted in different places on different chains

Thus, the same phenotype may result from changes at several different loci in the molecule or from multiple changes at a single locus, i e , a particular trait (e g , a given electrophoretic mobility) can result from an amino substitution located at several different positions in the α chain of hemoglobin, and the same trait can result from the same or different substitutions at several positions in the β chain, likewise the same trait can result from the substitution of different amino acids at a single position in one of the chains The latter phenomenon is referred to as *homoallelic mutation*, the former as *heteroallelic mutation*

In a broader sense, the study of hemoglobinopathies and other mutations provides excellent examples of how the confluence of genetics (both molecular and classical) and cellular processes at the molecular level becomes apparent in higher organisms Five genes determine the hemoglobin chains (α, β, γ, δ, and ϵ) The primary sequence of all but the ϵ chain has been determined, thus, the evolution of proteins can be discerned by noting homologous regions on the various chains This provides insights into meiotic processes which result in gene duplications, unequal crossing over, deletions, linkages, etc Moreover, since the amino acid sequences of the products are known, the relationship between structural alteration (amino acid replacement) and changes in function are revealed

The regulation of genetic processes can be studied by, for example, following the change from fetal hemoglobin (which possesses 2 α chains and 2 γ chains) to adult hemoglobin (which possesses 2 α chains and 2 β chains), this conversion is a result of the suppression of the γ gene and the activation of the β gene In sickle cell anemia, an abnormal hemoglobin is produced which causes red blood cells to assume a sickled or crescent shape This sickling trait is manifest at normal levels of blood oxygen tension in homozygotes, in whom it gives rise to a characteristic syndrome which proves fatal, in the heterozygote, this sickling trait is manifest only at low levels of blood oxygen tension The abnormality in the hemoglobin molecule that is responsible for sickle cell disease involves a change in only one amino acid (glutamic acid to valine) within each of the two beta chains of the globin tetramer It is noteworthy that this single amino acid substitution —which produces the aberrant hemoglobin— gives rise to a striking constellation of clinical signs, i e , splenomegaly jaundice, swelling of joints, cardiac enlargement bone deformities (kyphosis, tower shaped skull sabre shins), etc These clinical signs are all a consequence of the body's attempt to compensate for the malfunction of the erythrocytes due to the abnormal hemoglobin

A proteinopathy such as sickle cell anemia is an example of a qualitative genetic effect By contrast, in the group of disorders known as the thalassemias, the genetic effect is quantitative, i e , it arises from the fact that gene mutations mandate certain amino acid substitutions which reduce the rate of protein (hemoglobin) synthesis

This type of clinical molecular physiological analysis is, of course, not restricted to the hemoglobins As we establish the primary structure of other biologically important proteins, such as the immunoglobulins, their functions also become amenable to study by this approach

ROLE OF MUTATIONS IN MOLECULAR DISEASE

The foregoing is not by any means a complete summary of all the interactions by which genes determine phenotypes, in fact the question of genetic mutation has not yet been considered explicitly A mutation which changes the primary structure of the DNA and hence its information content--can be a single base change (point mutation) or can involve the deletion or addition of many bases Since these changes occur in the DNA, they are transmitted from generation to generation However, their expression is a function of all the foregoing considerations and the pattern of inheritance (i e , the pedigree) Thus, X-linked dominant alleles are expressed in males and females X-linked recessive alleles are carried by both males and females but expressed only in males Autosomal (nonsex-linked) dominant genes find expression in males and females whether homozygous or heterozygous for those alleles Finally, an autosomal recessive trait is expressed in both males (or females) only when it exists in the homozygous condition, however, heterozygotes are carriers of the trait Detailed examinations of heterozygotes have revealed certain biochemical abnormalities which do not interfere with normal body function Such findings provide the basis for genetic counciling by making it possible to identify heterozygote carriers of disease traits which, in the homozygote, would be deleterious

The frequency of mutations forms a continuum Rare mutations such as alcaptonuria occur with a frequency of approximately 10^{-6} Alcaptonuria is generally incompatible with life The mutation leading to blood type O has a frequency of 10^{-1} The defects in the enzymes which synthesize blood group antigens A and B are quite obviously compatible with life The requirement of all humans for ascorbic acid is an example of a mutation with a frequency of 100% Thus, mutations are not necessarily deleterious

Mutations need not always be expressed "Silent" mutations arise from base changes in DNA which, when translated during protein biosynthesis, yield the same or a functionally related amino acid (The degeneracy of the genetic code accommodates such events) Thus, a heritable change in DNA need not be expressed in its polypeptide product Alternatively, a mutation may be expressed structurally in a protein but may be "silent" in a functional sense, i e , the mutant protein may possess one or more amino acids which do not alter biological activity This type of structural polymorphism is exemplified by the transferrins (iron-binding plasma proteins) and haptoglobins (hemoglobin-binding plasma proteins)

A mutation can affect the rate of synthesis, and hence the intracellular concentration of proteins Quantitatively, this could lead to the absence or to the overproduction of an enzyme with deleterious metabolic consequences in either case Malfunction in this instance would not be a result of loss of activity due to change in the primary structure of the enzyme, but rather to a mutation in a control gene (such phenomena have not yet been demonstrated in humans)

Most expressed mutations result in loss of activity for the corresponding polypeptide The ensuing metabolic consequences depend on the role that this polypeptide fulfills in the cell as well as the interactions previously discussed The ramifications of an enzyme defect can be demonstrated by considering a generalized cell (Fig 1 126)

Absence or abnormal function of the permease, P_w, precludes the entry of metabolite, W, into the cell Limitation of this metabolite slows down the reaction sequence shown in the interior of the cell Although dietary intake or tissue synthesis of W may be substantial, the defective transport protein (permease) makes for starvation in the midst of plenty

The amino acid transport systems of the various tissues have the following common characteristics

1) The L-isomers of the amino acids are transported more efficiently than the D-isomers
2) Amino acid uptake occurs against an electrochemical gradient
3) Sodium, along with membrane-bound Na^+-K^+ ATPases, is involved in the translocation mechanisms
4) Concentrative uptake exhibits saturation kinetics, competitive and noncompetitive inhibition
5) The transport processes for the individual amino acids are functions of
   a) the length and composition of the aliphatic side chains (R groups)
   b) the anionic and cationic properties of the R groups
   c) the position of the NH_2 groups, α or β

The various amino acids are translocated across cell membranes via group specific transport systems To date, systems for the following groups of amino acids have been identified

dicarboxylic amino acids, dibasic amino

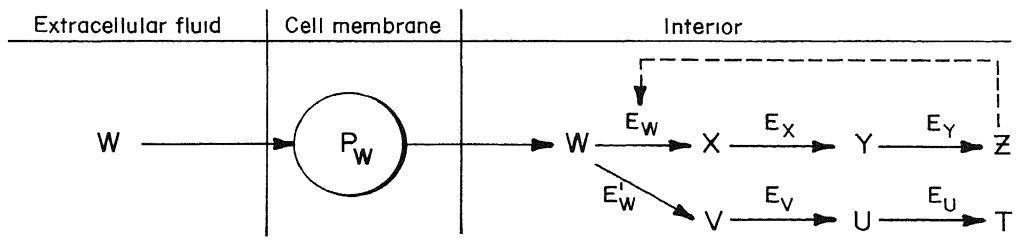


Fig 1 126 Diagram to illustrate the locations at which a mutation may alter metabolic pathways (see text) W X Y Z substrates of a major metabolic pathway V U T substrates of a minor (alternate) pathway P_W permease or transport system for metabolite W E_W E_X E_Y sequential enzymes in the major pathway E_W E_V E_U sequential enzymes in the minor pathway

acids, imino acids and glycine, neutral aliphatic and aromatic amino acids, and the β-amino acids

DEFECTS OF AMINO ACID TRANSPORT

The complexities of the genetic regulation of the amino acid transport systems become apparent in certain molecular diseases. Cystinuria is a disease in which the molecular lesion occurs in the transport system for cystine and dibasic amino acids (lysine, arginine, and ornithine), so that the renal tubular reabsorption is depressed. Consequently these amino acids are found in abnormally high quantity in the urine. The most distressing manifestation of this syndrome is the formation of cystine-containing renal calculi.

When this transport system is studied in several tissues obtained from cystinuric patients it is found that there are differences in its function in the gut, kidney, and leukocytes. The gut transport system loses its ability for concentrative uptake of its substrates (lysine, arginine, ornithine, and cystine). In the kidney, the situation is more complicated since there are other systems which transport one or more of these amino acids. In addition to the system which transports all four substrates, there is another system which transports only cystine, and a third system which transports only the dibasic amino acids. Renal clearance studies on cystinuric patients indicate that the former system is the site of the lesion and that the latter two systems are spared. In leukocytes there is no single system for all four substrates, but the cystine transport system is separate from the dibasic amino acid system. Moreover, both systems function normally in cystinuric patients. Thus, as noted earlier, the same gene may have different phenotypic expressions in different tissues.

The different phenotypic expressions of cystinuria are due to genetic variations, i.e., there is multiple allelism with respect to the relevant group specific permease system. The number of possible allelic combinations determine the number of possible genotypes. The genotypes can be ordered into classes on the basis of their phenotypic expression. The different (polymorphic) forms of this syndrome can thus be explained on the basis of genetic heterogeneity. Although these considerations apply to other molecular diseases, it should be emphasized that there are many possible combinations of genetic interactions and abnormal phenotypes (expressed in variations of protein structure and function) for which no prototype has been presented in this chapter.

Cystinuria is a member of the group of diseases known as aminoaciduria. Pathological aminoaciduria is characterized by an increased renal clearance of amino acids which is reflected in the urinary nitrogen content. The pattern of amino acid excretion in this group of diseases reflects the characteristics of the involved transport system and the site of the molecular lesion. A mutation which leads to a change in a carrier molecule, results in the abnormal urinary excretion of a specific amino acid (or group of related amino acids). If the molecular lesion occurs in the energy generating system, or the associated ion exchange system, a general aminoaciduria results. When the energy-generating system is involved, other substances not normally found in the urine will also be present. If the delivery of an amino acid present in the glomerular filtrate exceeds the reabsorptive capacity of its tubular transport system, it will spill over into the urine. Such a saturation aminoaciduria occurs when there is an accumulation of the amino acid in the plasma as a result of excessive intake, overproduction and/or defective metabolism.

Combined (multiple) aminoaciduria results from shared transport systems. An increased filtered load of one amino acid leads to competition for the reabsorptive carrier with the other amino acids of the group sharing the transport system. Thus, the former amino acid is excreted because its carrier is saturated and the others are also excreted because they cannot compete effectively for the saturated carrier.

Molecular lesions involving intracellular enzymes lead to precursor accumulation (Fig. 1.126). For example, substance W is not adequately metabolized because of a lesion involving enzyme, E_w. Accordingly the plasma concentration of W and, ultimately, the rate of urinary excretion of W may increase. Alternatively, if the molecular lesion involves enzyme, E_x, both W and X will accumulate with metabolic and excretory consequences depending on the nature of W and X. The disease entity, homocystinuria, is an example of this latter type of disorder in which the enzymatic lesion leads to accumulation of methionine and homocystine.

Accumulation of a precursor, W, may lead to

enhanced utilization of an alternate pathway (see Fig. 1.126, pathway $W \rightarrow V \rightarrow U \rightarrow T$). Increasing concentrations of potentially toxic metabolites along such an alternate pathway—the major metabolic route being impaired—may interfere with the normal cell function. Phenylketonuria (PKU) is an example of this class of disorder. In PKU the absence of phenylalanine hydroxylase leads by an alternate metabolic pathway to the abnormally enhanced production of phenylpyruvic acid, phenylacetic acid, and phenyllactic acid—all of which achieve toxic concentrations. Another consequence of an enzymatic lesion along a major metabolic pathway (see Fig. 1.126, $W \rightarrow X \rightarrow Y \rightarrow Z$) may result in deficit of Z, a physiologically active agent in the body. Goitrous cretinism, due to impairment of one or more enzymes necessary for the synthesis of thyroxin, exemplifies this class of disorders.

Finally, a molecular lesion may destroy feedback control by modifying an allosteric enzyme, so that it is no longer responsive to its inhibitor, Z, thus leading to overproduction of Z.

Space does not permit a more detailed consideration of the aminoacidopathies and proteinopathies. It should be emphasized, therefore, that there is an enormous amount of data available for study in this rapidly expanding area of inquiry, and thus the few examples given above are meant only to serve as signposts to the interested reader.

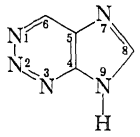
6 / PHYSIOLOGY OF PURINES AND PYRIMIDINES

This chapter deals with the biosynthesis of purine and pyrimidine bases, their incorporation into nucleic acids, breakdown of these acids, and the further catabolism of the liberated base derivatives Attention will be drawn to disorders of purines and pyrimidines in man, particularly those for which the etiology is understood in terms of enzymatic deficiencies and errors in the metabolic pathways

Many other important aspects of the physiology of purines, pyrimidines, and their derivatives not immediately relevant to this chapter, will only be mentioned briefly here These bases, covalently attached to various sugar-phosphates, represent "high energy" compounds One group, the acyclic nucleotides yield upon hydrolysis of the phosphate ester bond a negative free energy of the order of 7 5 kcal /mole This energy is available through coupled enzymatic reactions to drive endergonic biological processes, adenosine triphosphate (ATP), for example, provides energy for such processes as muscle contraction, biosynthesis of proteins, nucleic acids, carbohydrates, and lipids Nucleotides are also important in group transfer reactions The base moiety of the nucleotide determines specificity, e g , uridine nucleotides are necessary in glycogen biosynthesis, cytidine nucleotides in phospholipid biosynthesis, adenosine nucleotides in protein biosynthesis, etc Another group, the cyclic nucleotides, exhibits in comparison to acyclic nucleotides an even greater free-energy of hydrolysis amounting to approximately 12 kcal /mole Adenosine-3′,5′-cyclic monophosphate participates as intracellular mediator in the action of a number of hormones and biogenic amines via activation of cyclic adenosine monophosphate (AMP)-dependent protein kinases Cyclic AMP is also a physiological regulator in the induction of enzyme synthesis in bacteria and in mammals Evidence is emerging to indicate that cyclic nucleotides other than cyclic AMP are involved in similar physiological processes In addition, nucleotides frequently are components of coenzymes, e g nicotinamide adenine dinucleotide (NAD) nicotinamide adenine dinucleotide phosphate (NADP) and flavin adenine dinucleotide (FAD) all of which are involved in redox reactions catalyzed by various dehydrogenases Finally a most important aspect of the physiology of purines and pyrimidines is their function while incorporated in nucleic acids in a highly specific arrangement, as the carrier of the genetic information of all living organisms

Chemically speaking adenine and guanine are derivatives of purine, the structure of which is shown in Figure 1 127*A* Cytosine, uracil, and thymine are derivatives of the pyrimidine nucleus (Fig 1 127*B*) The bases when bonded to the sugar moieties ribose (Fig 1 127*C*) or deoxyribose (Fig 1 127*D*) via glycosidic linkage are referred to as nucleosides The nitrogen of the purine in position 9 or of the pyrimidine in position 1 is attached to the carbon 1′ in the sugar moiety Phosphorylation of a nucleoside in the 5′ position results in the formation of a nucleotide (nucleoside monophosphate), the monomeric unit of nucleic acid We speak of the polymer of ribonucleotides as ribonucleic acid (RNA) and of deoxyribonucleotides as deoxyribonucleic acid (DNA), depending on the nature of the sugar residue; polymerization occurs via ester formation between the 5′-phosphate of one nucleotide and the 3′-hydroxy group of another, etc While cytosine, adenine, and guanine are found in both RNA and DNA, uracil occurs essentially only in RNA and thymine in DNA

Table 1 12 lists the official nomenclature and abbreviations for the most frequent bases and derivatives encountered Trace amounts of unusual bases occur in nucleic acids, particularly in transfer-RNA (t-RNA), these include hypoxanthine, and methylated bases such as 5-methyl-cytosine, 6-methylamino purine, 1-methylguanine, or 2-methylamino-6-hydroxy-



Purine

a

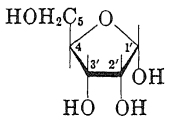
Adenine = 6-Amino-
Guanine = 2-Amino-6-Hydroxy-
Hypoxanthine = 6-Hydroxy-
Xanthine = 2,6-Dihydroxy-



Pyrimidine

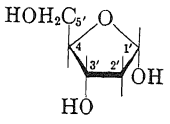
b

Cytosine = 4-Amino-2-Hydroxy-
Uracil = 2,4 Dihydroxy
Thymine = 2,4 Dihydroxy-5-Methyl-



Ribose

c



Deoxyribose

d

Fig 1 127 Structure of purines pyrimidines and the carbohydrate moieties present in nucleosides and nucleotides

Table 1 12
Nomenclature of common nucleotides[a]

Base	Nucleotides	Nucleotides (as Salts)	Abbreviations
Adenine	Adenosine	Adenosine-5′-phosphate or 5′-adenylate[b] Adenosine 3′,5′-*cyc* phosphate or *cyc* 3′,5′-adenylate	5′-AMP[b], ADO-5′-P, pA *cyc* 3′,5′ AMP, ADO 3′,5′-P
	Deoxyadenosine	Deoxyadenosine-5′-phosphate or 5′-deoxyadenylate	5′-dAMP[b], deADO-5′-P, pdA
Guanine	Guanosine	Guanosine-5′-phosphate or 5′-guanylate[b]	5′-GMP[b], GUO-5′-P, pG
	Deoxyguanosine	Deoxyguanosine-5′-phosphate or 5′-deoxyguanylate	5′-dGMP[b], deGUO-5′-P, pdG
Cytosine	Cytidine	Cytidine-5′-phosphate or 5′-cytidylate[b]	5′-CMP[b], CYD-5′-P, pC
	Deoxycytidine	Deoxycytidine-5′-phosphate or 5′-deoxycytidylate[b]	5′-dCMP[b], deCYD-5′-P, pdC
Uracil	Uridine	Uridine-5′-phosphate or 5′-uridylate[b]	5′-UMP[b], URD-5′-P, pU
Thymine	Thymidine	Thymidine-5′-phosphate or 5′-thymidylate[b]	5′-(d)TMP[c], THD-5′-P, pT
Hypoxanthine	Inosine	Inosine-5′-phosphate or 5′-inosinate[b]	5′-IMP[b], INO-5′-P, pI
Xanthine	Xanthosine	Xanthosine-5′-phosphate or 5′-xanthylate[b]	5′-XMP[b], XAO-5″-P, pX

[a] Abbreviations are in accordance with the IUPAC-IUB Tentative Rules (*J Biol Chem* 241 52, 1966)
[b] Omission of the 5′-designation is common in this series (e g , adenylate or AMP)
[c] 5′-Thymidylate, the deoxy derivative, is usually abbreviated TMP

purine Many of these unusual bases are more prevalent in viruses and bacteria than in higher organisms

Biosynthesis

The biosynthetic pathway of pyrimidines and purines has been worked out in great detail Generally, the approach was to administer to the experimental animal possible precursors, isotopically labeled, and then to determine the position of incorporation of label in the desired product

BIOSYNTHESIS OF PYRIMIDINES

The in vivo synthesis of uridine-5′-monophosphate, the key intermediate for pyrimidine

nucleotides and deoxynucleotides, entails the formation of a substituted pyrimidine ring (orotic acid) which is subsequently attached to a sugar monophosphate

The various steps leading to the synthesis of uridine-5'-monophosphate are outlined in Figure 1 128 In the first step, carbamoyl phosphate synthetase catalyzes the synthesis of carbamoyl phosphate from glutamine, bicarbonate, and ATP This activated phosphate ester donates its carbamoyl moiety for acylation of the α-amino group of aspartic acid to yield ureidosuccinic acid—a step catalyzed by aspartate transcarbamylase (ATCase) Ureidosuccinic acid undergoes ring closure and oxidation to orotic acid in two subsequent steps Orotidine 5'-monophosphate pyrophosphorylase catalyzes formation of the glycosidic linkage between orotic acid and phosphoribosyl pyrophosphate (PRPP) Irreversible decarboxylation catalyzed

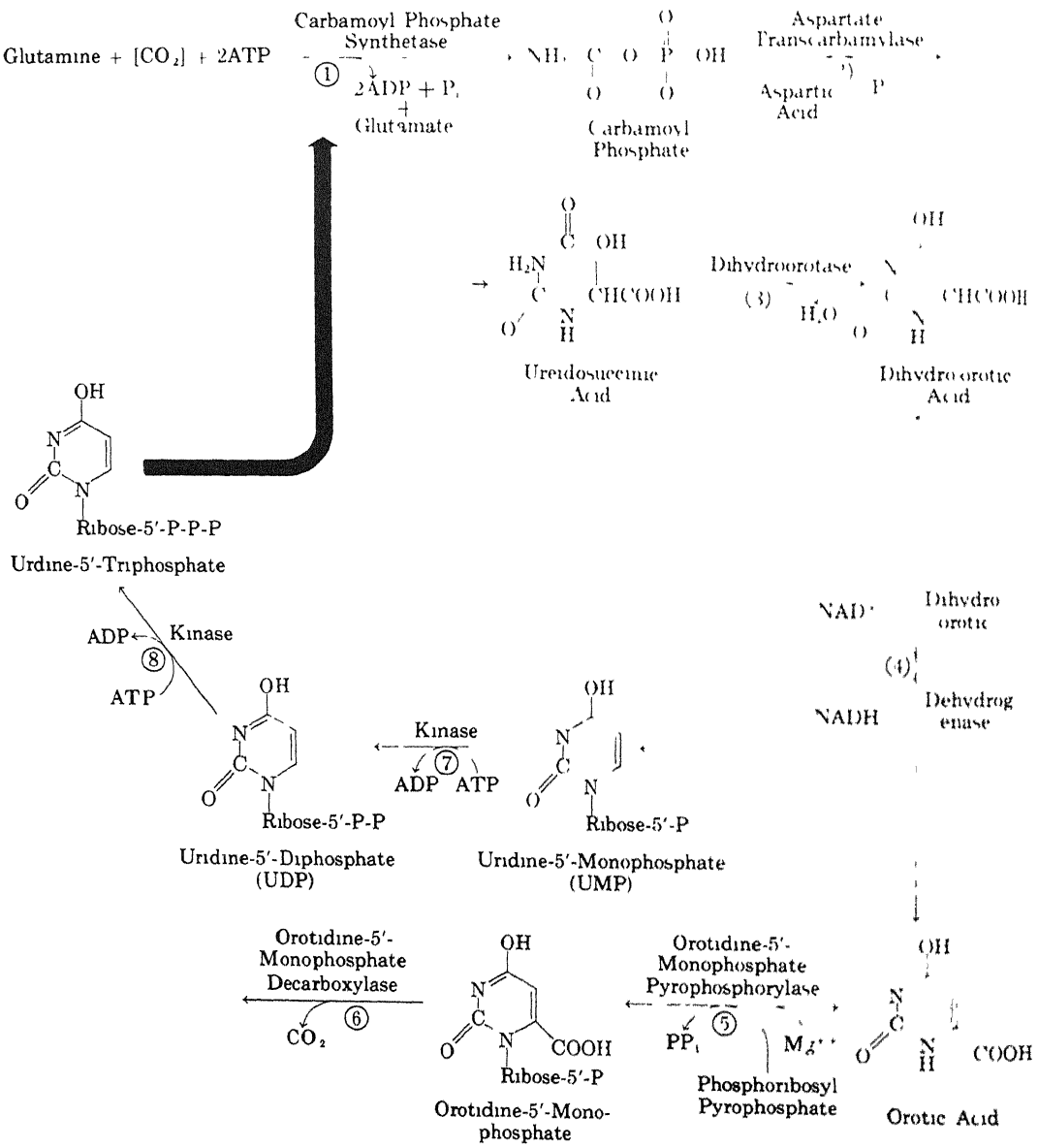


Fig 1 128 Biosynthetic pathway of uridine ribonucleotide Circled numbers indicate individual, known steps in the pathway Heavy arrow indicates feedback inhibition

by orotidine-5′-monophosphate decarboxylase, yields uridine monophosphate (UMP) With the synthesis of UMP, which is readily phosphorylated to the di- and triphosphate, a branching point in the biosynthetic pathway of the pyrimidine nucleotides has been reached On the one hand, part of the uridine nucleotide pool is incorporated into RNA and on the other, part is converted via uridine triphosphate (UTP) to cytidine triphosphate (CTP) (Fig 1 129*a*), the last step in this conversion involves the replacement of the 4-hydroxy group by an amino group, for which a purine nucleotide [guanosine triphosphate (GTP)] is required as cofactor

In addition, the UMP pool serves as starting material for the biosynthesis of deoxyribonucleotides It should be noted that theoretically deoxyribonucleotides could result either from the addition of a pyrimidine moiety to deoxyribose, or conversely, from the reduction of the sugar residue of the pyrimidine nucleotide, experiments in which radioactive ribonucleotides are incorporated intact into DNA favor the latter possibility In contrast to ribonucleotides, the elucidation of the biosynthetic pathway of deoxyribonucleotides in mammals has met with considerable difficulties Though the research in this area is in a state of flux, it is already apparent that the reductase systems are subject to stringent regulatory control dCDP seems to be synthesized in bacterial and mammalian systems via the route UMP → UDP → UTP → CTP → CDP → dCDP (see Fig 1 129*a*)

Studies in *Escherichia coli* would suggest that the remaining pyrimidine deoxyribonucleotide [deoxythymidine monophosphate (dTMP)] is synthesized via phosphorylation of UMP to UDP and reduction of the ribose moiety to deoxyribose This reduction requires several enzymes and cofactors, including two reductase enzymes, (B_1 and B_2), a sulfhydryl protein (thioredoxin), its reductase, and NADPH Finally, dUDP is dephosphorylated to dUMP, which in turn is methylated in position 5 by N^5,N^{10}-methylene-tetrahydrofolic acid in the presence of thymidylate synthetase to yield dTMP (Fig 1 129*b*) An apparently analogous pathway for the synthesis of dTMP is found in Novikoff hepatoma cells, in this mammalian system the conversion of UDP to dUDP likewise requires a reductase enzyme, two enzymes (S and S_b) which can be replaced by thioredoxin and its reductase, and NADPH, among other enzymes (Beck, 1969)

Having outlined the biosynthetic pathway for pyrimidine nucleotides we now consider its regulatory control Generally speaking, there are two types of feedback control inhibition and repression In the first case, a metabolic intermediate, usually an end product, inhibits one or more of the enzymes involved in its own synthesis, most frequently at an early step This is a short-term, rapid control, which prevents accumulation of unused product In the case of repression, the end product, usually a small molecule, decreases the synthesis of one or more enzymes involved in its biosynthetic path by acting at the DNA level, this control assures that proper enzyme concentrations are maintained in the cell In comparison to feedback inhibition, repression is a fine and long-lasting control because it is a response to a continuing situation rather than to momentary fluctuations Although these feedback mechanisms have been elucidated using microorganisms, there is evidence emerging which suggests that they are also operative in mammalian cells Extensive work has led in recent years to a considerable broadening of our knowledge of many aspects of the control mechanism of pyrimidine synthesis Originally, attention was focused on aspartate transcarbamylase (ATCase) (step 2 in Fig 1 128), since this enzyme had been shown to be subject to feedback inhibition in *E coli* (Gerhart and Pardee, 1962) ATCase from rat liver and from Ehrlich ascites tumor cells were reported to be sensitive to feedback inhibition However, later studies on purified liver preparations did not show sensitivity to inhibitors, nor did the enzyme when isolated from erythrocytes or leukocytes—even though in the latter study wide ranges of pH, and of substrate and inhibitor concentrations were tried (Curcie and Donachie, 1964, Prager, et al , 1967) The consensus at present is that ATCase does not play a regulatory role in mammalian systems, and attention has shifted to the first step in the pathway of pyrimidine synthesis, which is catalyzed by carbamoyl phosphate synthetase II (CPSase II) This enzyme, isolated from fetal rat liver, requires glutamine as nitrogen donor, and is subject to inhibition by UTP This latter finding, together with the fact that carbamoyl phosphate synthetase II exhibits a low turnover

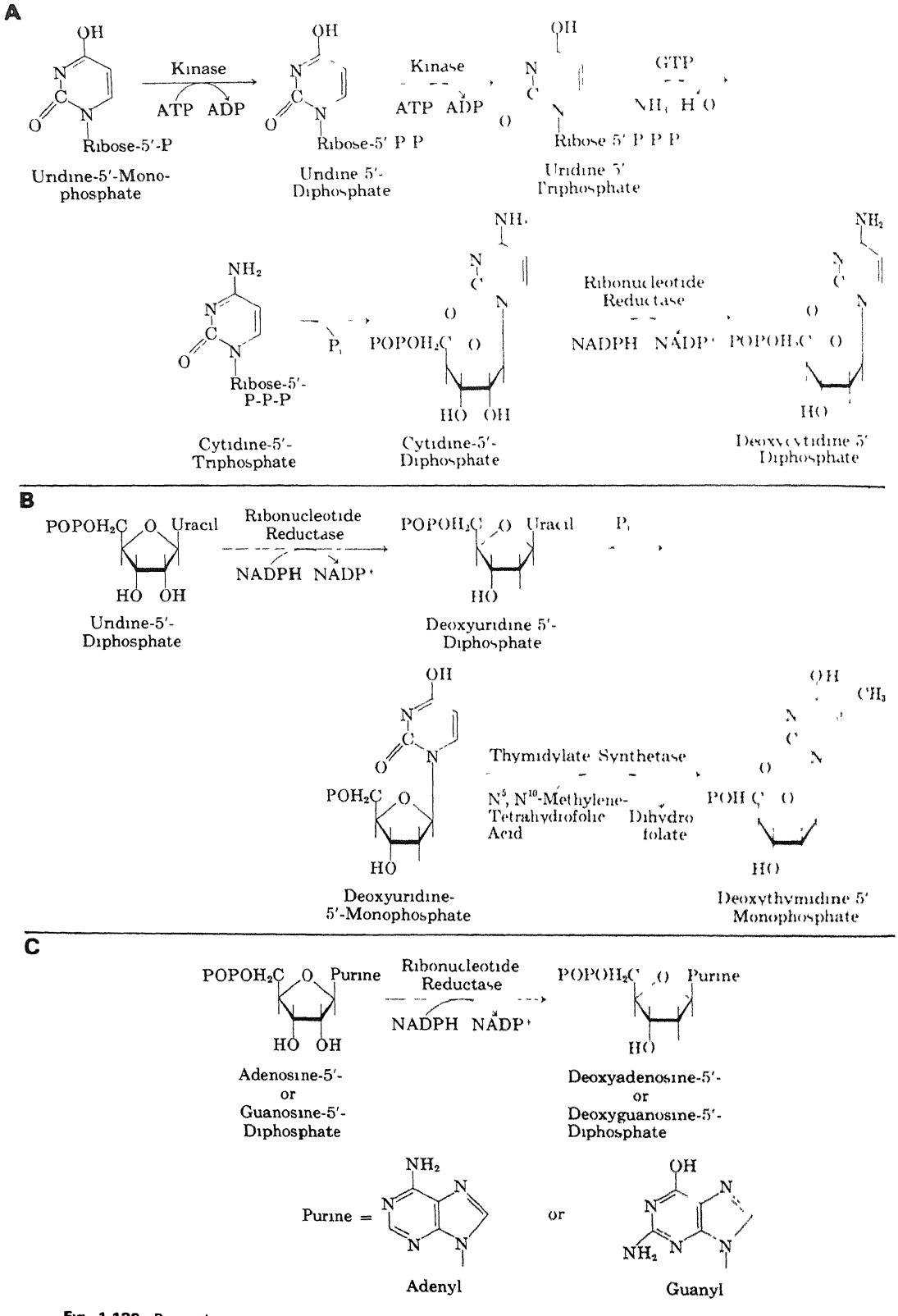


Fig 1.129 Biosynthetic pathways of cytidine ribonucleotide pyrimidine deoxyribonucleotides and purine deoxy ribonucleotides

rate, suggests that this enzyme may play an important regulatory role in the pyrimidine biosynthetic pathway (Hager and Jones, 1967)

It is worth mentioning that the liver of adult rats containd carbamoyl phosphate synthetase activity which requires ammonia as nitrogen donor and *N*-acetyl-L-glutamate as cofactor This carbamoyl phosphate synthetase is not sensitive to pyrimidine feedback inhibition, and evidence indicates that it is important in the urea cycle rather than in the synthetic pathway of pyrimidine From regenerating liver, in which both urea and pyrimidine synthesis are occurring simultaneously, carbamoyl phosphate synthetase activity has been separated into CPSase I (requiring ammonia) and CPSase II (requiring glutamine) (Nakanishi et al , 1968)

A metabolic intermediate can be involved in the regulation of a biosynthetic pathway not only through feedback inhibition but also by induction of increased levels of enzymes necessary to turn the intermediate into product Studies on human diploid cells in culture in which addition of dihydroorotic acid gives rise to increased levels of orotidine-5′-monophosphate decarboxylase (step 6 in Fig 1 128), and possibly also of the enzymes catalyzing the preceding two steps, may serve as an example (Pinsky and Krooth, 1967a, b)

BIOSYNTHESIS OF PURINES

In contrast to the pyrimidine ring, which attaches covalently to the ribose phosphate after its synthesis has been completed, de novo synthesis of the purine skeleton occurs on the ribose phosphate residue Studies with labeled precursors revealed the atoms of the purine skeleton to be derived from the following sources (Fig 1 130)

The first step in the biosynthesis of purine (see Fig 1 131) is the phosphoribosyl pyrophosphate amidotransferase-catalyzed amination of phosphoribosyl pyrophosphate (PRPP), which is an important intermediate and has been implicated as a regulatory factor in the biosynthesis of all nucleotides (recall that it appears also in step 5 of pyrimidine biosynthesis, see Fig 1 128) This step utilizes the amide nitrogen of glutamine, and involves an inversion at C-1 of the ribose moiety yielding 5′-phosphoribosyl-1′-amine This amine is then conjugated with glycine in an ATP-dependent reaction, and the resulting glycinamide ribonucleotide is formylated and subsequently amidated by transference of the amide nitrogen from glutamine The fifth step entails an ATP-dependent dehydration with concomitant ring closure to yield the imidazole portion of the purine The next two steps involve the carboxylation of 5-aminoimidazole ribonucleotide followed by amidation, the nitrogen being derived from the α-amino group of aspartic acid Into this intermediate, 5-aminoimidazole-4-carboxamide ribonucleotide, the last carbon atom is introduced by formylation and then, in a final step, dehydration leading to ring closure completes the synthesis of the hypoxanthine ribonucleotide, referred to as inosinic acid

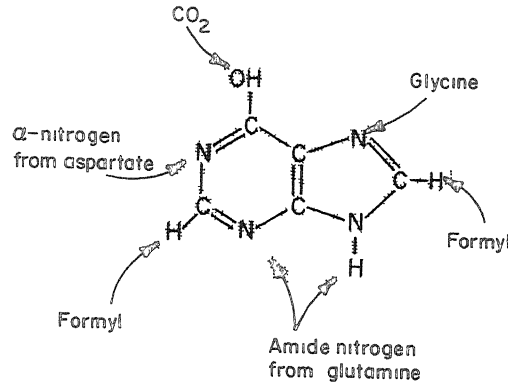


Fig 1 130 Source of atoms in the purine ring

With the synthesis of inosinic acid a branch point in the biosynthetic pathway has been reached (see Fig 1 132) On the one hand, amination of this intermediate in a two-step process which is GTP-dependent, results in adenosine-5′-monophosphate (AMP), on the other, likewise in a two-step process, inosinic acid is oxidized and then, in an ATP-dependent step, aminated to give guanosine-5′-monophosphate (GMP) Synthesis of purine deoxyribonucleotides is catalyzed in in vitro experiments by the same enzymes which reduce pyrimidine nucleosides Control is again stringent and complicated The pathway is outlined in Figure 1 129*c*

Purine biosynthesis is controlled by feedback inhibition at the first committed step, which is the glutamine-dependent amidation of PRPP catalyzed by PRPP-amidotransferase (step 1 in Fig 1 131) (Wyngaarden and Ashton, 1959) Studies of the partially purified enzyme from mammalian tissue showed that AMP, ADP, ATP, GMP, GDP, and inosine monophosphate (IMP) all can act as inhibitors with varying

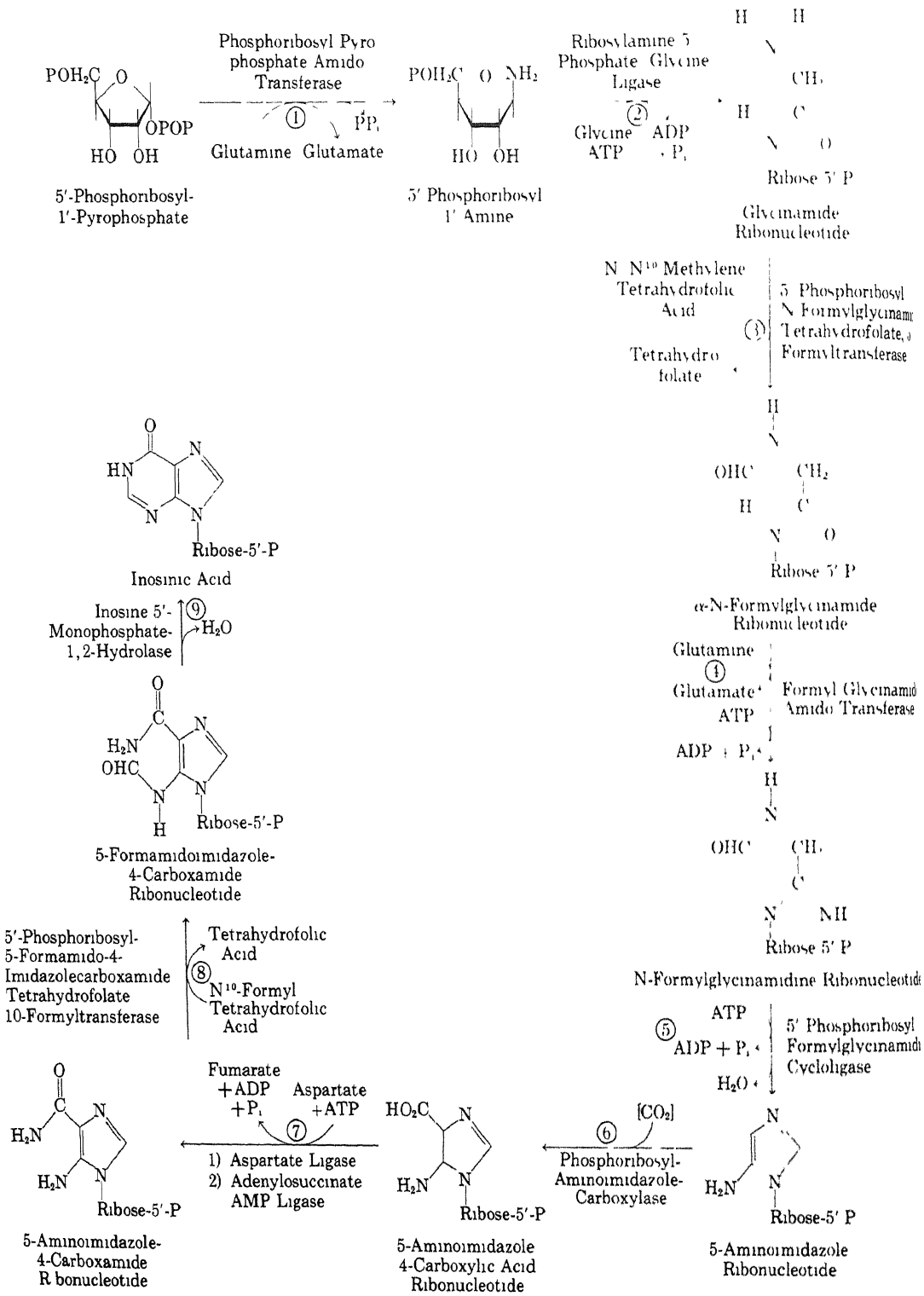


Fig 1 131 Biosynthetic pathway of inosinic acid

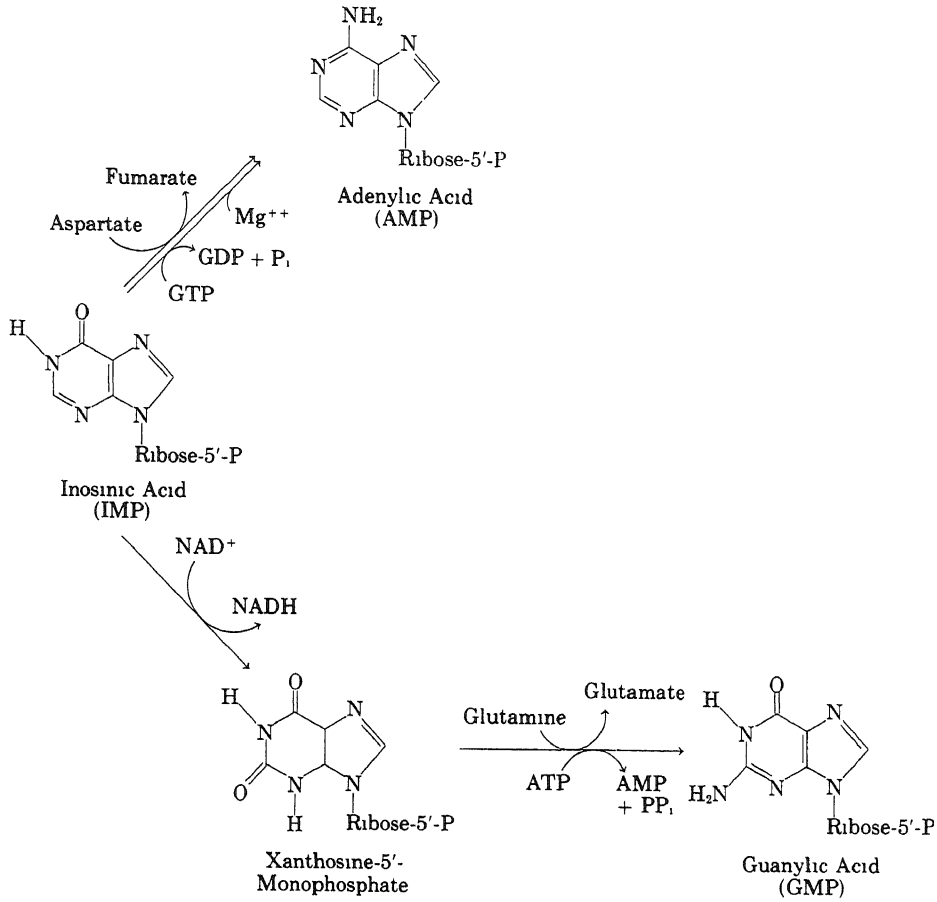


Fig 1 132 Conversion of IMP to AMP and GMP

degrees of efficacy (Caskey et al , 1964) In no case were purine nucleosides or free bases effective as inhibitors of the purified enzyme On the other hand, inhibition in mammalian tumor cells of de novo purine synthesis is obtained not only with nucleotides but also with purine bases and nucleosides It is most likely, in view of the enzyme studies mentioned above, that purine bases and nucleosides are converted to nucleotides within the cell before they can exert their inhibitory effects Pyrimidine bases, which are likewise in cells readily converted to their corresponding nucleosides and nucleotides, are incapable of inhibiting PRPP-amidotransferase, this reveals the inhibition of the transferase as feedback rather than depletion of the substrate PRPP in the course of nucleotide formation Further, the conversion of inosinic acid to AMP requires GTP for energy, and the conversion to GMP requires ATP We may view these reciprocal requirements as a kind of regulatory control In addition, in bacteria, the enzyme involved in the oxidation of inosinic acid is subject to feedback inhibition by GMP (Fig 1.133), (Magasanik, 1958), and one of the enzymes necessary for AMP formation has been shown to be inhibited by a number of purine nucleotides (Wyngaarden and Greenland, 1963)

REUTILIZATION PATHWAYS OF PYRIMIDINES AND PURINES

In addition to engaging in de novo pyrimidine and purine synthesis, many cell types are capable of utilizing preformed bases, derived essentially from breakdown of nucleic acids or from ingested food, for nucleotide formation (Murray et al , 1970) Although a great deal of similarity appears to exist between the types of enzymes involved in the reutilization of pyrimidines and purines, to date more emphasis has

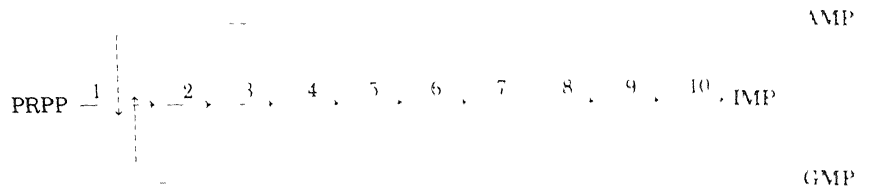


Fig 1 133 Schematic representation of main control points in purine biosynthesis. Circled numbers correspond to steps in the biosynthetic pathway shown in Figure 1 131. Dashed lines indicate regulatory pathways.

been placed on the purines, because of their wider use as therapeutic agents. It was only in recent years, however, that the lack of certain reutilization enzymes was recognized as a cause of serious disorders.

The most important enzymes for both pyrimidine and purine reutilization are *a*) nucleoside phosphorylases, *b*) nucleoside kinases, and *c*) phosphoribosyl transferases (Fig. 1 134). Nucleoside phosphorylases catalyze the reversible reaction

$$\text{nucleoside} + P_i \rightleftharpoons \text{base} + \text{ribose-1-P}$$

These enzymes, possessing broad specificity, have been found in many tissues, such as spleen, liver, and Ehrlich ascites tumor cells. Nucleoside kinases phosphorylate nucleosides to yield nucleotides. Most studies have been performed using adenosine kinase, which is widely distributed in mammalian tissue. Adenosine kinase is fairly specific for adenosine, although it can react with certain adenosine analogs. Phosphoribosyl transferases catalyze the formation of the glycosidic linkage between base and PRPP, resulting in nucleoside monophosphate and PP_i. In mammalian tissue two enzymes have been well characterized to date—one which reacts with adenine and the other with hypoxanthine, guanine, and xanthine.

In summary, preformed bases can be converted to nucleotides either by formation of nucleoside and subsequent phosphorylation, or by direct condensation with PRPP. The specificity of the reutilization pathway enzymes and their mode of control assure that there is no excess of one base or nucleotide at a given time and that de novo synthesis can be kept at its lowest level for maximum efficiency. These pathways also assure that if exogenous purine or pyrimidine is available, it can be utilized to the greatest benefit of the organism. In addition to the above enzymes involved in the reutilization of preformed bases, many additional enzymes exist, likewise carefully controlled,

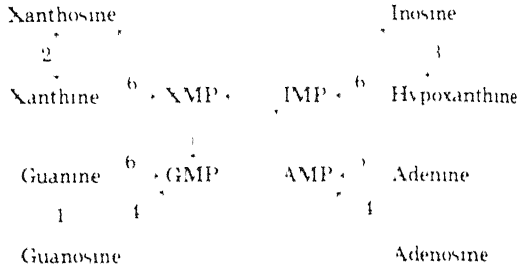


Fig 1 134 Pathways of purine reutilization. The most important enzymes for purine reutilization are: Nucleoside phosphorylases (1, [illegible]); Nucleoside kinases; 4, Adenine phosphoribosyl transferase; 5, Hypoxanthine-guanine phosphoribosyl transferase; 6.

which catalyze the conversion of one pyrimidine to another, and, similarly, of one purine to another.

Incorporation of Purines and Pyrimidines into Nucleic Acids, Their Orientation in Nucleic Acids, and Nucleic Acid Breakdown

The deoxyribonucleoside triphosphates of adenine, guanine, thymine, and cytosine are the substrates for enzyme-catalyzed polymerization which results in deoxyribonucleic acid (DNA), as depicted in Figure 1 135. Similarly, ribonucleoside triphosphates of adenine, guanine, uracil, and cytosine are polymerized by mammalian DNA-dependent RNA polymerase, yielding ribonucleic acid (RNA). The resulting nucleic acid molecules are the genetic information carriers of the living organism. The information transfer reactions, including duplication of DNA, transcription, which is the synthesis of messenger RNA on a DNA template, and translation of the messenger RNA code, are processes in which the formation and cleavage of highly specific hydrogen bonds between purines and pyrimidines play a crucial role. Hydrogen bonds are the result of an associative interaction between an acidic, hydrogen-

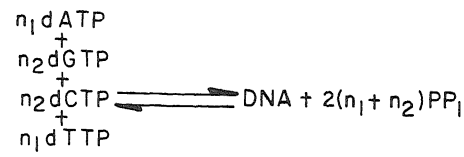


Fig 1 135 Schematic representation of enzymatic DNA synthesis

bearing atom (donor), and an electronegative group (acceptor), such as tertiary nitrogen or carbonyl oxygen

For some time the specificity of purine-pyrimidine pairing (guanine-cytosine, adenine-thymine) was thought to reside in the stereochemical restraints imposed by the helical DNA structure, however, x-ray studies of crystals as well as spectroscopic studies in solution of base derivatives themselves have indicated that the cooperativity of all noncovalent bond interactions, including the hydrophobic bond, ionic bond, but particularly the hydrogen bond, is sufficient to bring about specific base-pairing (Voet and Rich, 1970) For example, several configurations are theoretically possible for the guanine-cytosine base pair (Fig 1 136), but on the basis of the model systems studied, it can be concluded that only one configuration (*b*) occurs This configuration, as opposed to others, derives added stability from the high number of favorable hydrogen bonds In fact, this pairing coincides with the Pauling-Corey modification of the Watson and Crick proposal in the DNA model The elucidation of the information contained in nucleic acids ranks among the great scientific achievements of our time, and the interested reader is referred to pertinent summaries (Taylor, ed , 1963, 1967, Frisch, ed , 1966 and 1969, Crick, 1966, Yanofsky, 1967)

The turnover rate of DNA and all the different types of RNA in mammalian cells is slow Although several enzymes capable of degrading DNA and RNA have been isolated and well characterized, our understanding of the in vivo breakdown of these molecules is at present rather limited

Among the known deoxyribonucleases attention should be drawn to two, both of which are endonucleases, but which differ in their mechanism of cleavage DNase I, which is effective at neutral pH, catalyzes cleavage of nucleotide bonds to form oligonucleotides possessing a free 3′-hydroxy group, but retaining the 5′-phosphate Another type of nuclease, DNase II, has an acidic pH optimum, and catalyzes intranucleotide cleavage to yield oligonucleotides with a 5′-hydroxy and a 3′-phosphate group Both types of deoxynuclease have been isolated from bacterial as well as mammalian sources (Laskowski, 1967) Similarly, microorganisms and mammalian tissue contain ribonucleases which possess different specificities Bovine pancreatic RNase, for example, a well-characterized protein whose complete structure has been elucidated (Wyckoff et al , 1968), is an endonuclease which catalyzes the cleavage of the 3′-phosphate ester as long as a pyrimidine base—rather than a purine—is attached at the 1′-position of the given ribose moiety It appears that DNA and RNA are first converted to oligonucleotides which in turn are catabolized to mononucleotides, possibly by nonspecific phosphodiesterases The 5′-nucleotidases and nucleoside phosphorylases, so far found in all tissues investigated, then complete hydrolysis to the free bases The further breakdown of purines and pyrimidines is discussed in the following sections

DEGRADATION OF PYRIMIDINES

Our present knowledge of the catabolic pathways of cytosine, uracil, and thymine is summarized in Figure 1 137 Much of this information is derived from work with microorganisms and enzymes isolated from mammalian liver A problem in the determination of these pathways has been the lability and short lifetime of some of the intermediates The first step in the degradative pathway of pyrimidines is a reduction to the dihydro derivative Hydrolytic ring scission of the amide bond between positions 3 and 4 gives rise to the corresponding *N*-carbamoyl-β-amino acid These compounds are deacylated to β-alanine in the case of uracil and β-aminoisobutyric acid in the case of thymine That these compounds are susceptible to further degradation by transamination to malonic- and methyl malonic-semialdehyde and eventual retutilization for branched-chain fatty acid synthesis is known from earlier work

DEGRADATION OF PURINES

Because one product of catabolism of purines is uric acid, a compound implicated in several human diseases, the degradative pathway of the purines has been worked out in greater detail than that of pyrimidines Purine catabolism

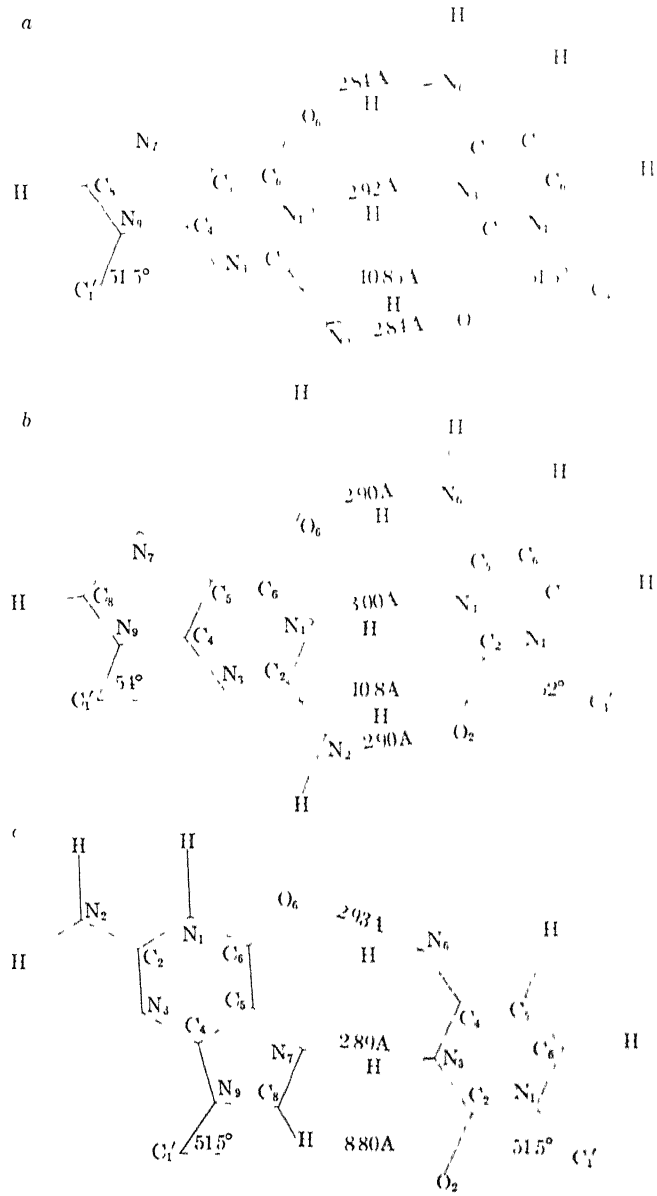


Fig 1 136 Some theoretically possible guanine-cytosine base pairs *a)* Watson Crick model *b)* Pauling Corey modification of Watson-Crick model *c)* model with two instead of three hydrogen bonds

is outlined in Figure 1 138 Guanine is deaminated by guanase to xanthine The analogous reaction is not possible with adenine since animal tissue lacks adenine deaminase, hence, in a three-step reaction adenylic acid or adenosine are likewise converted to xanthine (Balis, 1967) The last of these steps is catalyzed by xanthine oxidase, which also oxidizes xanthine to uric acid (Fig 1 138, steps 3 and 4) Uric acid is the excretory product in man, some primates, reptiles, fowl, insects other than *Diptera*, and in the Dalmatian dog (which cannot reabsorb uric acid) In the ureotelic mammals, primeval vertebrates, marine elasmobranchs, and teleosts, uric acid is further hydrolyzed by uricase to allantoin, a more soluble excretory product Allantoin, in turn, is further degraded to allantoic acid by allantoinase, and then to urea and glyoxylic acid by allantoicase. In the lower forms of life, bacteria and marine inver-

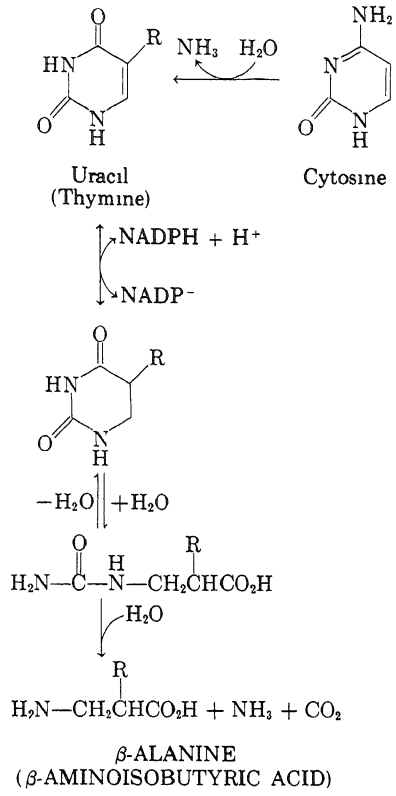


Fig 1 137 Degradative pathway of pyrimidine ribonucleotides Uracil R = H thymine R = CH_3

tebrates, urea is not excreted, but instead is metabolized to carbon dioxide and ammonia It is tempting to speculate that higher animals have lost uricase in the course of evolution, leaving them with the disadvantageous excretion of the slightly soluble uric acid (Gutman, 1965)

Uric acid is not the major pathway of nitrogen excretion in man Instead, the ammonia nitrogen of most amino acids, the major nitrogen source, is shunted into the urea cycle The major amount of uric acid is excreted by the kidney, although uricolytic bacteria present in the small intestine degrade the acid further, as was revealed by administration of labeled uric acid and recovery of label in expired carbon dioxide, as well as in urinary uric acid A greater amount of extrarenal elimination of uric acid appears to occur in patients who are hyperuricemic; this finding has been thought to be a cause and effect relationship, but the increased extrarenal elimination may also be due to undetected impaired kidney function

Disorders of Purine and Pyrimidine Metabolism

GOUT

Primary gout is a genetically determined disorder of purine metabolism characterized biochemically by hyperuricemia Hyperuricemia may be asymptomatic for long periods, in gout it usually becomes clinically significant through recurrent attacks of acute arthritis The arthritic stage rarely begins before puberty in males or before menopause in females Renal calculi can occur at any time, but tophaceous deposits usually appear after the development of acute arthritis Gouty patients may overproduce and overexcrete uric acid, or they may excrete normal amounts In certain cases production may be normal, but defective renal function may result in hyperuricemia which in turn may cause gout (renal gout) Secondary gout is acquired as a complication of hyperuricemia associated mainly with hematological malignancies, or in drug-induced states

One of the major problems in gout is the nature of the triggering mechanism of the acute attack Uric acid is rather insoluble in acidic solutions, although alkalinity increases its solubility At blood pH (7 44) it is present as the sodium salt, which is likewise insoluble but frequently forms supersaturated solutions The events leading to initial crystallization of sodium urate from supersaturated fluid are poorly understood, but once crystals form they are phagocytized by numerous polymorphonuclear cells in the synovial fluid It has been recently shown that the phagocytosis of urate crystals, but not of diamond dust particles, results in the elaboration of a chemotactic substance by the polymorphonuclear cell (Phelps, 1969) It now seems likely that the primary action of colchicine in aborting gouty inflammation is its inhibition of motility of the polymorphonuclear cells by interference with the formation of protoplasmic gels (microtubules) (Malawista, 1968)

The most common definition for hyperuricemia is statistical, i e , any value over two standard deviations from the normal mean is defined as high The upper limit of normal has been found to be 6 3 to 6 4 mg /100 ml in males, and 5 8 mg /100 ml in females (by the uricase method) These values depend somewhat on the technique used for measurement, colorimetric methods giving higher values than enzymatic ones Another widely used measure-

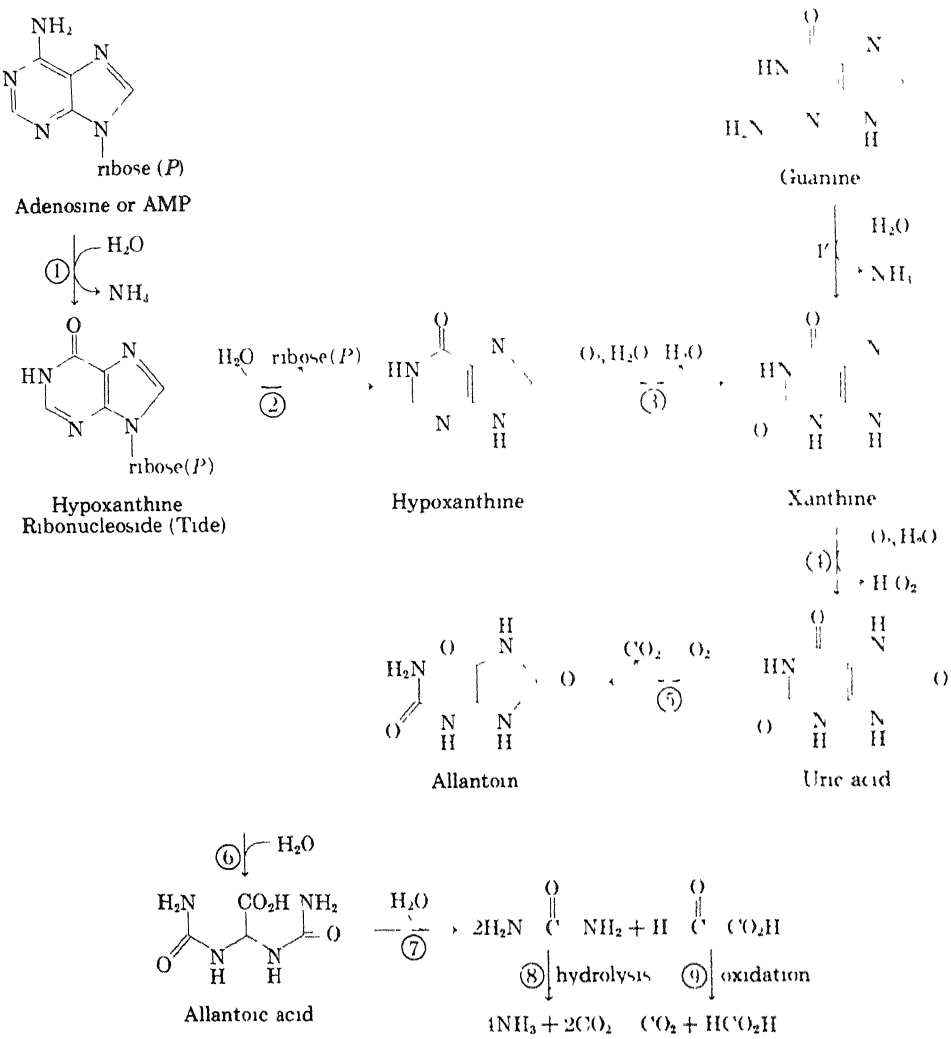


Fig 1 138 Degradative pathway of purine ribonucleotides Circled numbers indicate steps see text

ment is that of the "miscible pool" of uric acid, determined by isotope dilution of urinary uric acid, after an injection of the labeled compound In normal males the pool contains about 1200 mg , with a turnover of approximately 700 mg /day The serum uric acid concentration in 98% of gouty patients was found to be above 7 mg /100 ml by the uricase method In these subjects the miscible pool was generally as high as 2000 to 4000 mg , and in cases of severe gout could reach 18,000 to 31,000 mg. (Wyngaarden, 1966) There is some speculation that urate may exist in bound form in plasma, and that a deficiency in binding may cause the excess (Sorenson, 1969). In this chapter the intention is to relate causes of gout as they are understood in terms of disorders in the basic biochemistry of the metabolic pathways of purines described above

Uric acid is formed from purines, which are liberated as a result of enzymatic degradation of tissue and dietary nucleoproteins and nucleotides, and is also formed by de novo purine synthesis. The immediate precursors of uric acid are hypoxanthine and xanthine, which are both acted upon by xanthine oxidase The rate of purine synthesis in normal and gouty patients is routinely studied by measuring incorporation of ^{15}N- or ^{14}C-glycine into uric acid In contrast to normal subjects, patients with primary gout show enhanced uric acid production. The pattern of incorporation of label into uric acid is characterized by a rapid and excessive rise, followed by a fairly rapid decline

This pattern is indicative of a "shunt pathway" for uric acid synthesis, i e , a major portion of the newly synthesized purine is diverted from incorporation into nucleic acids and instead is catabolized to uric acid

Most forms of primary gout (metabolic gout) are a result of overproduction of uric acid caused by accelerated de novo purine synthesis, which may or may not be accompanied by overexcretion It is to the cause of the increased synthesis that we address ourselves below

In 1964, an X-linked hereditary disease, now known as the Lesch-Nyhan syndrome, was first reported Its clinical features are developmental regardation, spastic cerebral palsy, choreoathetosis, and self-mutilating, aggressive behavior (Bland, 1968) The patients have markedly increased serum uric acid levels, usually in the range of 15 mg /100 ml , and excretion of uric acid is so high (in terms of milligrams uric acid per kilograms body weight) that these children would be classified as overexcretors even if they were adults Incorporation studies with ^{14}C-glycine indicate an overproduction up to 20 times control, which is about 10 times the uric acid level in gout From other studies it is known that hypoxanthine is normally recycled to a much greater extent than xanthine, and therefore the ratio xanthine hypoxanthine is always greater than one In the Lesch-Nyhan syndrome, more hypoxanthine than xanthine is excreted This fact indicates that hypoxanthine is probably the major precursor of uric acid in the syndrome The explanation of this finding is that all patients exhibiting the syndrome totally lack the enzyme hypoxanthine-guanine phosphoribosyl transferase (PRT), which catalyzes the transfer of the ribose-5′-phosphate moiety of 5′-phosphoribosyl-1′-pyrophosphate to the free purine bases, hypoxanthine and guanine (Fig 1 134, step 6) The absence of the enzyme is especially notable in the brain, where activity is normally high The conclusion which can be drawn is that lack of this enzyme evokes a detrimental increase in purine synthesis There are at least two possible mechanisms *1)* a decrease in GMP formation could result in diminished feedback inhibition of phosphoribosyl amidotransferase, *2)* a decreased utilization of PRPP may increase its availability for de novo synthesis (both Fig 1 131, step 1)

Because of the similarity of the Lesch-Nyhan syndrome to gout, it was natural to assay for PRT in gouty patients Seegmiller and his colleagues carried out extensive work both in vitro, on cultured human fibroblasts, and in vivo (Kelley et al , 1969) They found a decrease in PRT activity in a substantial number of gouty subjects The decrease is the same within families, but the absolute amount of decrease is different between families All patients with a lack or defect of PRT produce excessive quantities of uric acid, with a reflection in the urinary excretion in most cases PRT-deficient patients show the "shunt" pattern of incorporation Isolated mutant enzymes show varying degrees of heat stability that are quantitatively different from normal enzyme In the course of the studies it was consistently found that the patients had an abnormally high level of adenine phosphoribosyl transferase (A-PRT) The significance of this observation has not been determined, although it is tempting to speculate that the lack of GMP derepresses this enzyme, since GMP can be converted to AMP through IMP Cultured fibroblasts from the enzyme-deficient subjects showed an increase in purine synthesis to four times normal, and in some cases an increase in free intracellular PRPP, probably due to decreased utilization Fibroblasts deficient in PRT also showed that amidotransferase activity could be *activated* by hypoxanthine or guanine, unlike the situation in which the nucleotides of these bases serve as inhibitors This activation could provide another possible mechanism for increased purine synthesis de novo, especially since intracellular nucleotide concentrations seem to be normal, so that feedback inhibition should not be diminished These studies on gouty overproducers of uric acid indicate that the enzyme PRT, formerly believed to be a rather unimportant "salvage" enzyme, may play an important role in control of purine metabolism Whether the loosening of this control causes increased synthesis by raising the concentrations of substrates (PRPP, hypoxanthine, and guanine), or by decreasing inhibition, or both, will most likely be determined in the near future

Most of Seegmiller's gouty patients produced quantities of uric acid comparable to those discussed above, but had normal PRT activity This fact demonstrates that primary gout is the phenotypic expression of a heterogeneous group of genetic abnormalities Other possible causes will now be discussed

Decreased utilization of PRPP has already been implicated, however, the erythrocytes of

some gouty patients show an increased capacity for the synthesis of this substrate (Hershko, et al , 1968) The overall effect of increased synthesis of PRPP is the same as its decreased utilization, both phenomena leading to accelerated de novo purine synthesis The increased PRPP synthesis is linked to enhanced levels of endogenous ribose-5'-P, its immediate precursor

Some experiments have demonstrated that excessive purine synthesis in gouty patients can be modulated by adenine, indicating that derangements of the regulatory mechanism are quantitative rather than qualitative Also, in vitro leukocyte studies showed that while ^{14}C-glycine incorporation was on the average 1 8 times higher in gouty than normal cells, hypoxanthine was incorporated to the same extent in both cell types, and still inhibited excessive purine synthesis In these examples feedback inhibition and PRT activity are both normal, however, Rosenbloom, Henderson, and Kelley (1968) did demonstrate two patients with normal PRT activity and a defect in feedback inhibition (Fig 1 131, step 1) Incorporation of ^{14}C-glycine by skin fibroblasts was less sensitive to exogenous adenine and hypoxanthine in these patients than in normal individuals (Rosenbloom et al , 1968)

It has been postulated that in gouty patients increased quantities of glutamine are available for the first step of purine biosynthesis in the liver due to decreased glutaminase activity This decreased activity could impair the renal formation of ammonia and result in unduly acid urine This postulate was advanced in an attempt to explain the development of uric acid stones in gouty patients and their rarity in nongouty hyperuricemic patients Earlier observations had indicated that uric acid stone formation (urolithiasis) resulted chiefly from hyperexcretion of uric acid, however, approximately three-quarters of gouty patients with stones excrete normal amounts of uric acid Although direct studies have failed to confirm a deficiency of glutaminase in renal tubules, there is a persistent undue acidity of the urine of patients with primary gout accompanied by an absence of early morning and postprandial alkaline tides This finding correlates with reduced ammonia formation, 25 to 30% less than expected in relation to the urinary pH It is this persistent excessive acidity of the urine which is currently thought to be the major factor in stone formation (Yu and Gutman, 1967)

In summary a deficiency of PRT activity can cause uric acid overproduction in gout, and while it is rather rare it provides a rational explanation for overproduction in these particular patients The enzymes of de novo purine biosynthesis are under strict regulatory control, however, occasionally a loosening of this control does occur PRT deficiency contributes to this occurrence Probably no one specific cause of hyperuricemia and gout may be postulated, but the various biochemical lesions leading to the disease have either been pinpointed or are under focused investigation Clinically gout and, in general, hyperuricemia are treated with allopurinol, a purine analog which effectively inhibits xanthine oxidase in vitro and in vivo, and which also appears to inhibit de novo purine synthesis The consistent reduction in uric acid synthesis in response to allopurinol is a result of both actions

XANTHINURIA

Xanthinuria is a rare disorder of purine metabolism characterized by excretion of xanthine as the major end-product of purine metabolism Oxypurine levels of the serum are elevated, and uric acid levels are low Because xanthine is the chief purine metabolite excreted, it was thought that there might be a lack of xanthine oxidase activity (Fig 1 138, steps 3 and 4) This fact has been confirmed by direct assay Hypoxanthine does not accumulate under these conditions because of its efficient reutilization Xanthine oxidase, a flavin protein containing molybdenum and iron as cofactors, is in mammals normally found in the liver, small intestinal mucosa and possibly bone marrow. The complete absence of xanthine oxidase in patients with xanthinuria provides a clear demonstration of a specific biochemical lesion as the cause of a disease, although for this particular disorder no broader clinical manifestations have been reported except for the occasional passing of a xanthine stone

HEREDITARY OROTIC ACIDURIA

Hereditary orotic aciduria is a rare disease characterized by retarded growth and development, megaloblastic anemia, and excessive ex-

cretion of orotic acid. The detection in urine of orotic acid, a known intermediate in pyrimidine biosynthesis, suggested that there might be a block in the biosynthetic pathway subsequent to orotic acid formation (Fig. 1.128). The first patient discovered with this abnormality was therefore treated with yeast pyrimidine nucleotides. His improvement led to assay for the enzymes of pyrimidine biosynthesis in samples of his cultured erythrocytes and those of his family. Low levels of both orotic acid pyrophosphorylase and orotidine monophosphate decarboxylase activities were found (Fig. 1.128, steps 5 and 6) (Smith et al., 1961). Increased incorporation of labeled precursor in these patients suggested accelerated de novo pyrimidine synthesis, probably due to release of end-product inhibition. This disease is a good example of how an hereditary disorder can be pinpointed and treated. Once the complete biosynthetic and catabolic pathways of a particular compound are known, the excretion of an intermediate metabolite strongly suggests that a step subsequent to its formation has been blocked. The lack of two enzymes, as noted in this case, might indicate an abnormality in the genetic regulatory mechanism. Treatment of orotic aciduria has been limited to replacement therapy with uridine (Becroft et al., 1969). Success of treatment depends on the efficiency of reutilization pathways for pyrimidines. The higher efficacy observed with uridine as compared to uracil suggests that the more important reutilization path in man is normally the two-step one involving the nucleoside phosphorylase and nucleoside kinase. A problem with uridine treatment has been that the blood-brain barrier may not be permeable to the nucleoside, causing continued pyrimidine starvation in the brain which may explain the intractable mental retardation. Treatment does reduce the excretion of orotic acid, however, so that the exogenous nucleoside re-establishes feedback control. The pyrimidine reutilization pathways, here useful in control of disease, may well be recognized to be essential in normal metabolism as has been found in the case of purines.

7 / NUCLEIC ACIDS AND GENE ACTIVITY

Nucleic acids are macromolecules present in all living cells They were first isolated from pus cells by Friedrich Miescher in 1869 Complete hydrolysis of nucleic acids yields a mixture of basic substances called purines and pyrimidines a sugar and phosphoric acid The nucleic acids are classified into two main categories, according to the sugar they contain Those containing D-ribose are called *ribonucleic acids* or RNA Those containing D-2-deoxyribose are called *deoxyribonucleic* acids or DNA The pyrimidines commonly found in nucleic acids are *uracil* (U), *thymine* (T), and *cytosine* (C), the major purines are *adenine* (A) and *guanine* (G)

Deoxyribonucleic Acid

The first serious consideration that DNA can carry genetic information arose from Avery's experiments (1944) on bacterial transformation Bacterial transformation may be defined as the transfer of genetic characteristics to a recipient cell by means of DNA isolated from a donor cell population The phenomenon of bacterial transformation was observed for the first time in *pneumococcus* All virulent pneumococci form smooth colonies (S-type) when cultured on agar, because of a distinctive polysaccharide capsule on the cell surface Avirulent strains of pneumococci do not have the ability to form the polysaccharide capsule when cultured on agar, and hence form rough colonies (R-type) Griffith made the observation in 1928 that the addition of heat-killed pneumococci of the S-type into a suspension of live avirulent pneumococci caused a small fraction of these bacteria to be transformed into pathogenic cells Subsequently Avery, McLeod, and McCarty (1944) in a now classical experiment showed that pneumococci of the R-type can be transformed in vitro into the S-type by means of DNA prepared from R-type cells Confirmation that nucleic acids are the carriers of genetic information came from studies with viruses From these studies it was shown that *viral genes* are also nucleic acids

STRUCTURE OF DNA

The Covalent Backbone of DNA The deoxyribonucleic acids are polynucleotides in which the fundamental units the nucleotides, are linked to one another with 3′–5′ phosphodiester bonds Thus the covalent backbone of DNA is made of alternating deoxyribose and phosphate groups To this backbone the purine and pyrimidine bases are attached like side chains (Fig 1 139) The structure of polynucleotides is often represented schematically as shown in Figure 1 140 It can be seen that a polynucleotide has a *3′-end* and a *5′ end* These ends are very important in establishing directions of synthesis of DNA or RNA, as well as in direction of reading of RNA during protein synthesis

A shorthand way of writing nucleic acid sequences is to symbolize the bases by their initial letter and the phosphate group by the letter p A p positioned to the left of the base designates the 5′-position, if p appears to the right of base designates the 3′ position Finally, the prefix d- symbolizes deoxyribonucleic acid sequences Thus, d pApTpCpCpT symbolizes a deoxypentanucleotide in which *a*) the 5′ end has a free phosphate group *b*) the four phosphodiester bonds are between 3′A—5′T, 3′T—5′C, 3′C—5′C, 3′C—5′T, and *c*) the 3′-end has a free —OH

Base Composition of DNA Application of quantitative chromatographic methods in the analysis of various DNA preparations enabled Chargaff (1951) to draw the following extremely important conclusions

1) The base composition of DNA varies from species to species

2) DNA preparations obtained from various tissues of a given species have the same base composition, which is characteristic for that species

3) The base composition of DNA of a species does not change with age or nutritional state

4) In every DNA examined the number of adenine residues was equal to the number of thymine residues, that is, A = T Similarly it

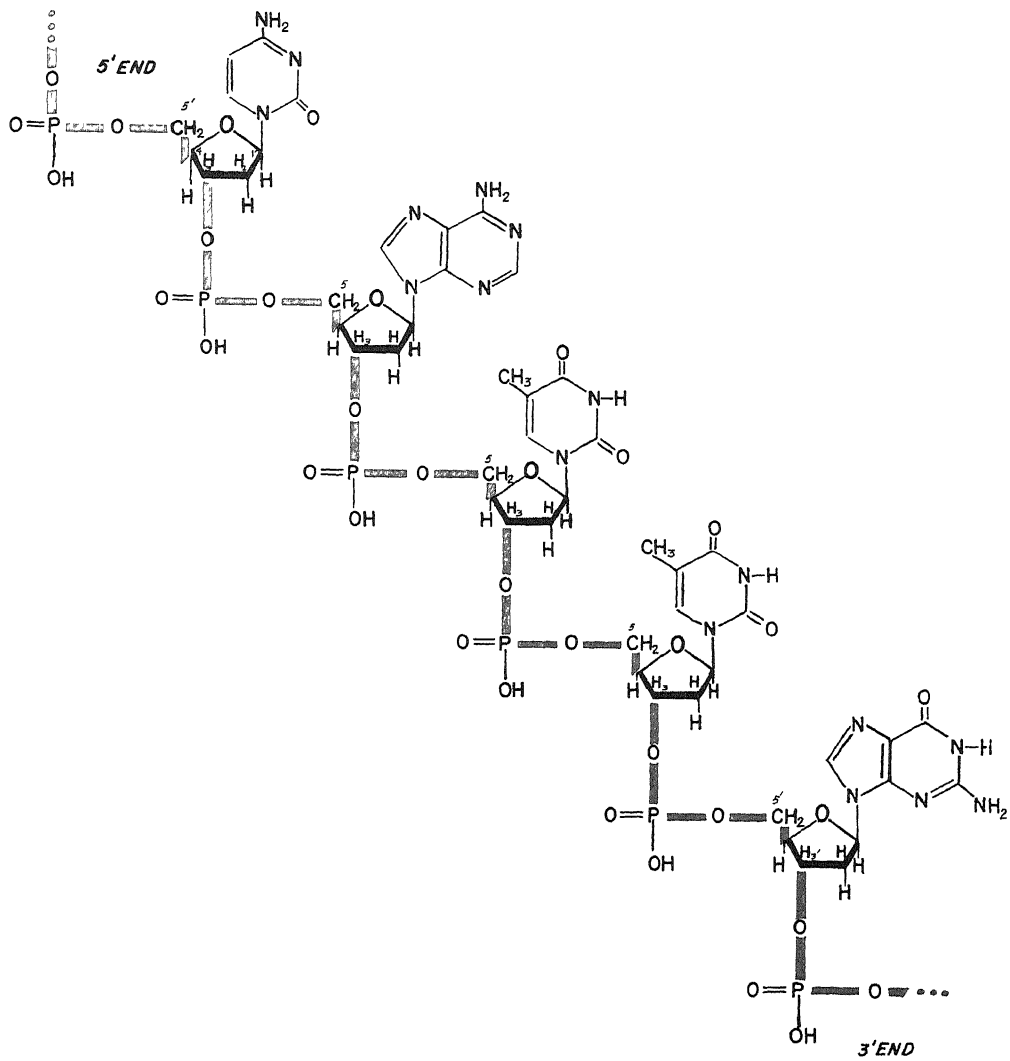


Fig 1 139 Structure of a portion of a DNA chain with hypothetical base sequence 5' to 3', cytosine-adenine-thymine-thymine-guanine

was found that G = C, and therefore A + G = C + T In other words the sum of the purine residues is equal to the sum of the pyrimidine residues These relationships apply to all double stranded DNA but not the single stranded ones

Three-Dimensional Structure of DNA. Application of x-ray crystallographic techniques to fibers of highly purified DNA enabled Wilkins and his colleagues to obtain important data on the three-dimensional structure of DNA DNA was found to possess two periodicities—a major one corresponding to a regular spacing every 3 4 Å along the fiber axis, and a secondary one every 34 Å (Fig 1 141) The crystallographic data together with the equivalence of bases in DNA described by Chargaff set up the stage for the understanding of the 3-D structure of DNA In 1953 Watson and Crick proposed a model for DNA which accounted for all the observed chemical and physical properties of DNA The characteristics of the Watson-Crick model are the following

1) DNA consists of two right-handed helical polynucleotide chains, coiled around the same axis to form a double helix

2) The two polynucleotide chains (strands) are antiparallel, in the sense that the 3'—5' phosphodiester bonds run in opposite directions

3) The bases of the chains are stacked on the

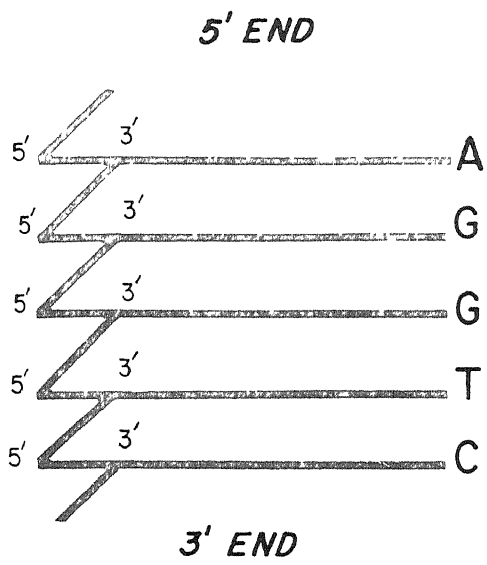


Fig 1 140 Schematic representation of a polynucleotide chain

inside of the double helix The planes of bases are parallel to each other and perpendicular to the long axis of the duplex

4) The bases of one chain are paired with hydrogen bonds with the bases of the other strand This pairing occurs between bases lying in the same plane Only certain base pairs can be accommodated, these are A-T and G-C (Fig 1 142) In other words A always pairs with T and G always pairs with C

5) The bases are stacked on top of one another and at a distance of 3 4 A There are exactly 10 bases per complete turn of the double helix (34 A)

6) The double helix has one major groove and a minor one (Fig 1 143)

Generally speaking the hydrophilic sugar residues and charged phosphate groups are on the outside of the DNA molecule whereas the hydrophobic bases are stacked within the double helix where they are shielded from water

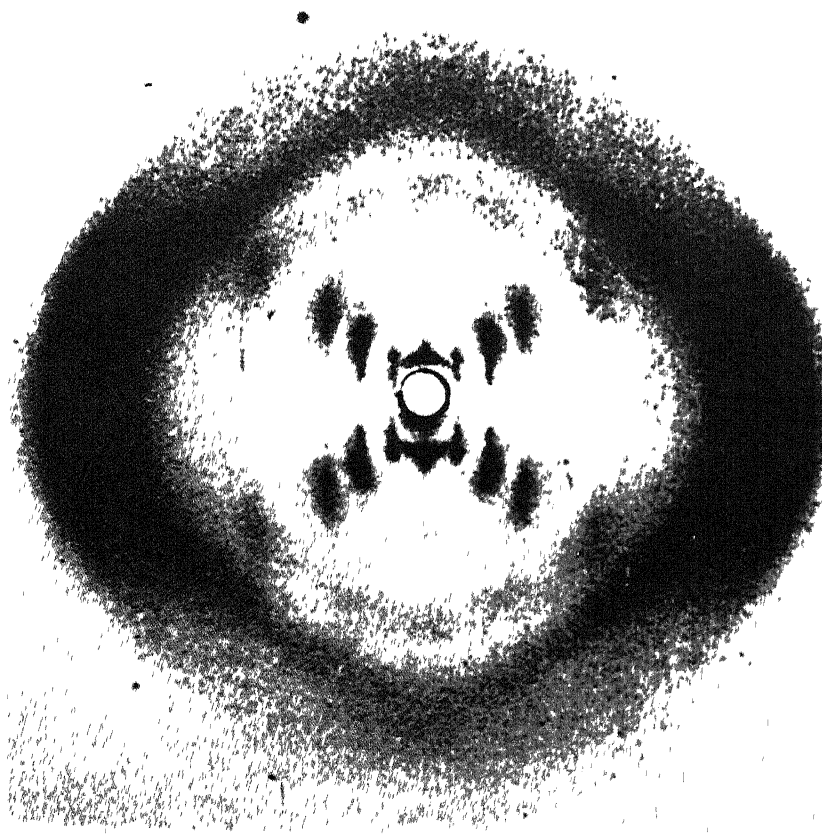


Fig 1 141 X-ray diffraction photograph of DNA The helical form of DNA is indicated by the pattern of x-ray reflections crossing in the middle From Franklin and Gosling, 1953

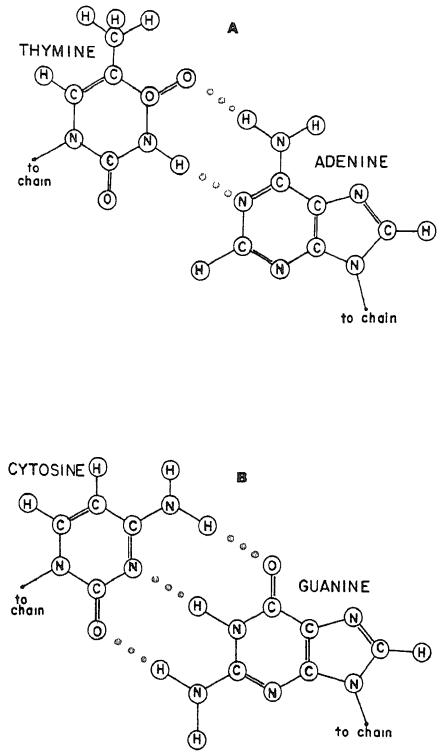


Fig 1 142 Base pairing in DNA through specific hydrogen bonding *a)* Thymine-adenine *b)* Cytosine-guanine

DNA carries the genetic information in the form of the base sequences This coded information is translated through certain mechanisms into amino acid sequences This means that the genetic information for the precise sequence of amino acids incorporated into proteins is stored in DNA in the form of base sequences

The complementary double helical structure of the Watson-Crick model of DNA suggests that the molecule can be accurately self-replicated Indeed it is evident that each strand can serve as a template for the synthesis of the other strand The complementarity of the DNA strands and the specific and accurate base pairing are the key features for the DNA-replication

Size and Shape of DNA Molecules. Intact DNA molecules are extremely large in size In fact native DNA molecules of bacteria are so large that they are easily broken by shear force, during their isolation The bacterial virus T2 contains one DNA molecule of MW = $1\,3 \times 10^8$ The single chromosome of *Escherichia coli* consists of a single DNA duplex MW = 2×10^9 having a contour length of almost 1 mm Eukaryotic cells contain several chromosomes and each chromosome may contain one or more extremely large DNA molecules Although very little is known about the size and shape of the DNA molecules of higher organisms they are believed to be much larger than those of bacteria Although, initially, the experimental evidence suggested that DNA molecules were *linear*, as the techniques of handling DNA

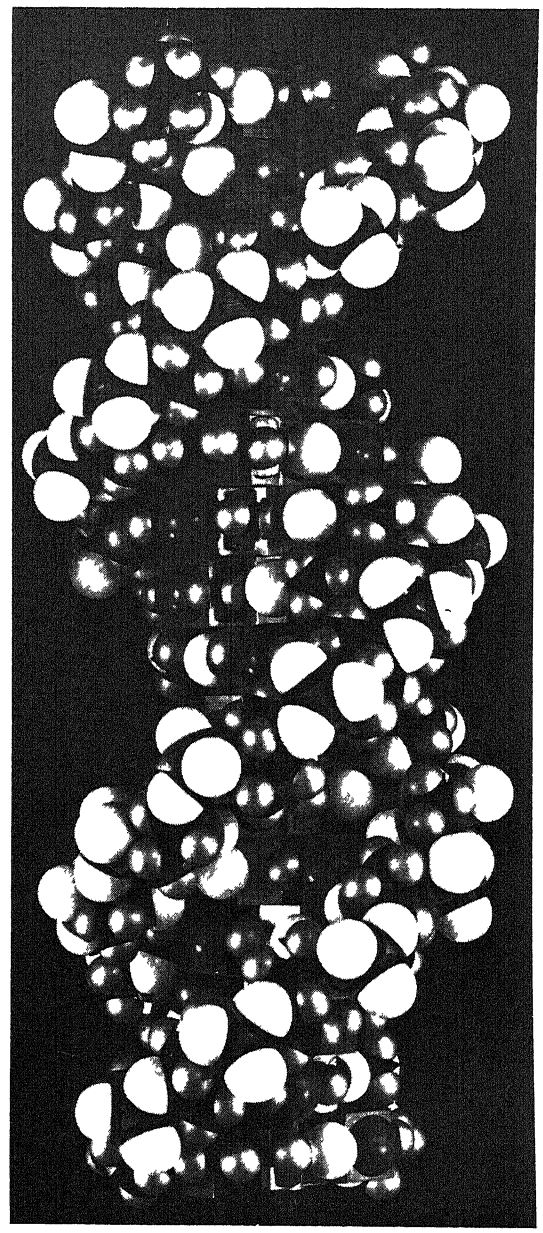


Fig 1 143 Space-filling model of DNA Courtesy of M H F Wilkins

preparations were refined the picture emerged that many DNA molecules are normally present in a *circular form*. DNA from λ phage has been found to be linear when isolated from the virus but circular when inside the host cell. This suggests that the linear and the circular forms are interconvertible. It is possible in this case that the circular form is assumed during replication of the phage DNA, whereas the linear form is necessary for the injection of DNA through the narrow tail of the phage into the host cell. Circular DNAs are found not only in viruses but also in bacteria. Thus the single chromosome of *E. coli* consists of a single circular DNA duplex. Finally, in certain viruses as for example the ϕ 174 the DNA is a circular single strand (Fig. 1.144).

THE GENETIC ORGANIZATION OF DNA

The relative positions of different genes on chromosomes can be determined by genetic experiments. The phenomenon of *genetic recombination* provides the basis for the *chromosome mapping* experiments. Genetic recombination occurs between homologous chromosomes which specifically attract each other and form pairs. Following pair formation, sometimes the chromosomes break at the same point and recombine crosswise (*crossover*). In bacteria crossover and genetic recombination

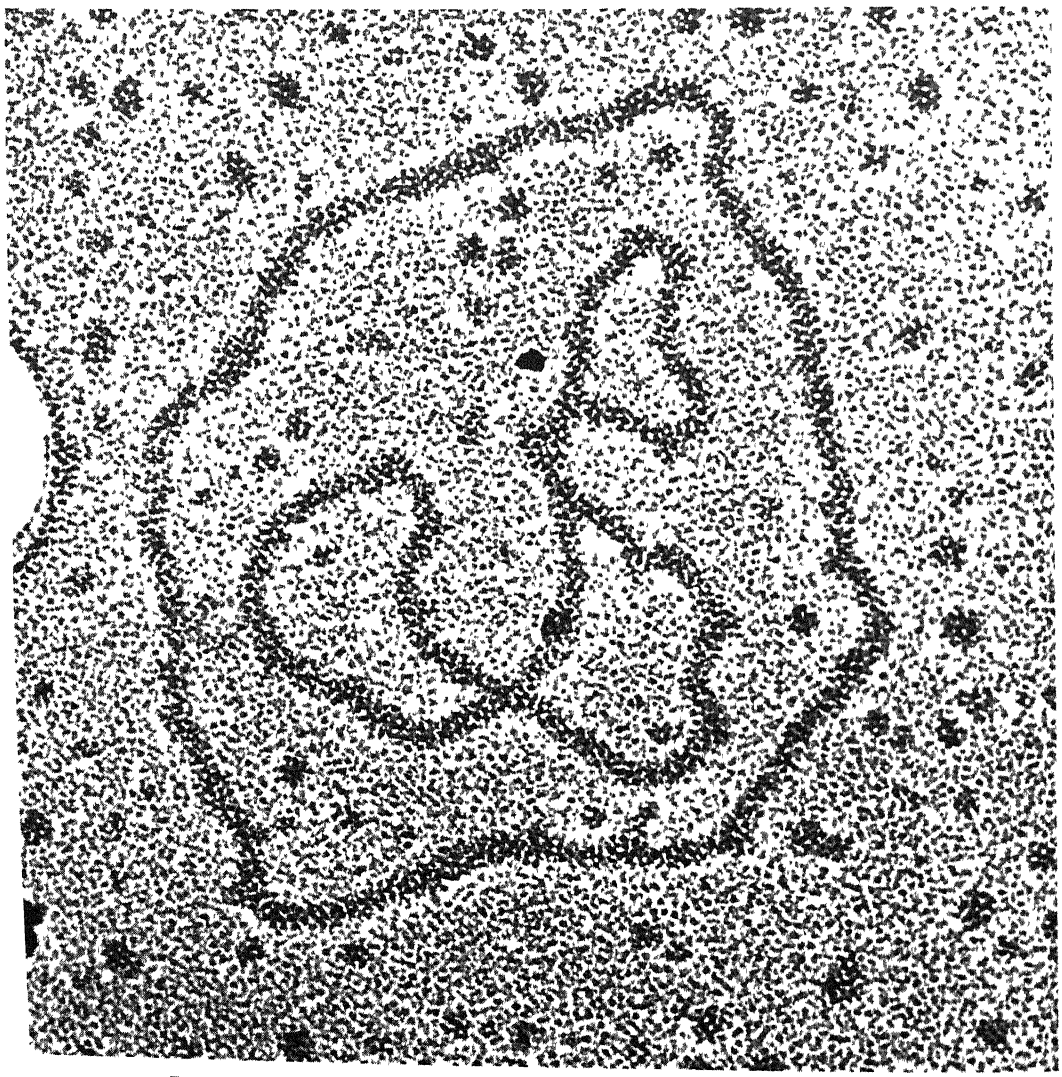


Fig. 1.144 Electron micrograph of DNA from ϕ×174 phage. From Lehninger 1970.

usually involve the intact chromosome of an acceptor cell and a chromosome fragment from a donor cell Transfer of the DNA fragment to the acceptor cell occurs, for example, during bacterial mating (*conjugation*), when a segment of the chromosome of one cell is transferred to the other, transfer also takes place during *transformation* (see Avery's (1944) experiment) The identity of the transferred and recombined genes can be deduced from the induced changes in the acceptor cell Consider for example R-type pneumococcus transformed in vitro into the S-type by means of DNA In this case some of the transferred genes control the formation of the characteristic polysaccharide capsule of the S-type pneumococcus

By employing techniques that permit analyses of these phenomena, various genes have been located on the chromosome, and detailed genetic maps for viruses and bacteria have been deduced These and other studies have provided the proof that one gene controls one protein The early "one gene - one enzyme" hypothesis thus became an established fact Since a protein may consist of more than one polypeptide chain, a more correct statement is "one gene - one polypeptide chain " For example, the protein tryptophan synthetase of *E coli* consists of four polypeptide chains, two of one type, called A, and two of another, called B The A and B polypeptides are coded for by two adjacent genes Usually but not always, genes coding for different polypeptide chains (subunits) of a protein are adjacent on a chromosome

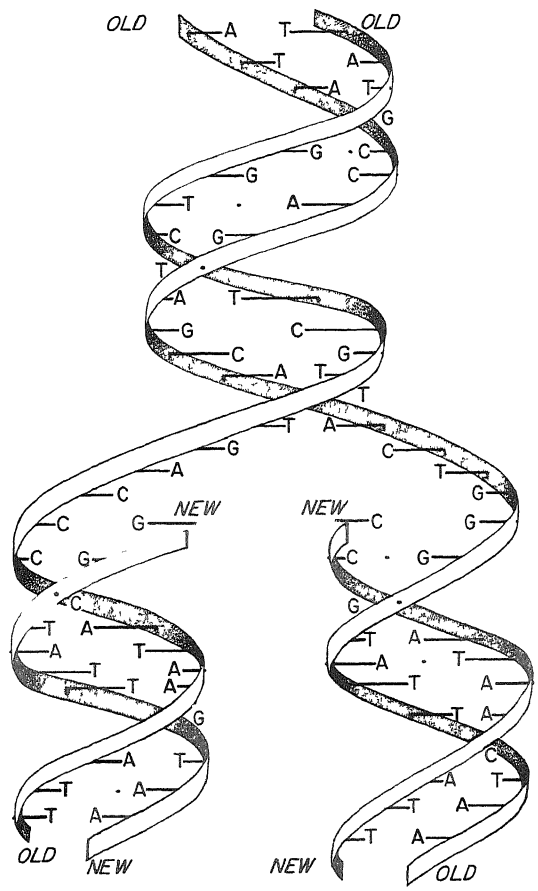


Fig 1 145 Semiconservative replication of DNA according to Watson-Crick (Watson 1970)

DNA REPLICATION

The Semiconservative Model of DNA-Replication Watson and Crick (1953) postulated that DNA can self-replicate and that the key to this process lies in the complementarity of the strands and the accuracy of the specific base pairing According to the proposed mechanism, the strands of the original duplex separate Each strand, as it separates, serves as a template for the synthesis of a new complementary strand The newly synthesized strand, together with the parental strand, forms a daughter DNA molecule (Fig 1 145) This process is called *semiconservative replication* Such a model of DNA replication was supported by the classical Meselson-Stahl (1958) experiment *E coli* bacteria grown in a medium containing the heavy isotope ^{15}N have denser DNA than when they are grown under normal conditions with ^{14}N The "heavy" DNA can be separated from the "light" DNA by equilibrium density gradient centrifugation in CsCl solution In the Messelson-Stahl experiment *E coli* cells grown in ^{15}N-medium were transferred to ^{14}N-medium One generation after the transfer, all the DNA was found to have hybrid density i e , between the densities of the "heavy" and the "light" DNA, after two generations, half of the DNA had hybrid density and half was light (Fig 1 146)

The Moving-Fork Mechanism Use of tritiated thymidine during replication results in tritium labeled DNA This DNA can be visualized by autoradiography (Fig 1 147) Autoradiographs of this kind revealed that the new strands of DNA are synthesized as the parental strands separate In the autoradiograph in Figure 1 147, one can observe the entire rep-

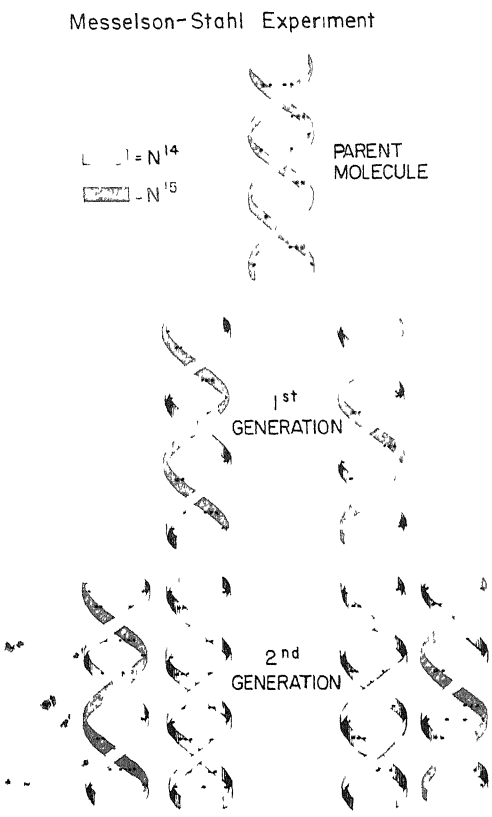


Fig 1 146 Replication of N^{15}-DNA (heavy) in the Messelson Stahl experiment In first generation each daughter molecule contains one heavy strand (N^{15}) and a light one (N^{14}) In second generation two DNA molecules are hybrids and two are light These results present solid evidence in favor of semiconservative replication

licating chromosome of *E coli* Studies of autoradiographs of DNA isolated at various stages during the replication process enabled Cairns (1963) to observe the following *a*) the DNA molecule appears as a continuous double stranded circle, *b*) the replication appears to be sequential from a fixed starting point, *c*) the synthesis of daughter strands is accomplished by a moving-fork mechanism, and *d*) behind the moving fork one can observe the two separated daughter duplexes, which are joined again at the initiation point The fork moves at a rate of 27μ min $^{-1}$ which corresponds to the addition of 10^5 nucleotides to each synthesized strand Studies in bacterial and viral systems on the duplication of certain genes, showed that replication of DNA always initiates at *starting points* uniquely placed relative to known genetic loci Replication proceeds sequentially in a unique direction relative to the starting point The starting points remain invariant over repeated cycles of replication It appears that the complex processes of DNA replication and DNA repair (which also requires synthesis of new material) involve several enzymes such as *DNA polymerase DNA-ligase* and specific *endonucleases* Recent information implicates a *membrane bound enzyme* as the principal enzyme of replication The function and role of all these enzymes in DNA-replication and DNA repair is considered below

DNA Polymerase DNA polymerase (A Kornberg 1969) catalyzes the addition of mononucleotide units to 3′ end of a DNA at a rate of 1000 nucleotides per minute per molecule of enzyme DNA polymerases have now been isolated from a variety of bacterial and animal cells All these enzymes catalyze the following reaction

$$n_1\text{dATP} + n_2\text{dGTP} + n_3\text{dCTP} + n_4\text{dTTP} \xrightleftharpoons{\text{DNA, Mg, Enzyme}} \text{DNA} \begin{bmatrix} (\text{dAMP})_{n_1} \\ (\text{dGMP})_{n_2} \\ (\text{dCMP})_{n_3} \\ (\text{dTMP})_{n_4} \end{bmatrix} + (n_1 + n_2 + n_3 + n_4)\text{PP}$$

$$\downarrow \text{Pyrophosphatase}$$

$$2(n_1 + n_2 + n_3 + n_4)\text{P}_i$$

In the reaction shown above for each phosphodiester bond synthesized, a pyrophosphate bond is broken and a pyrophosphate group is released This reaction has a slightly positive free energy. The reaction is driven to completion because splitting of the high energy bond of pyrophosphate by the pyrophosphatase releases additional energy. The reaction requires preformed DNA whose function is considered below, and takes place only in the presence of all four deoxyribonucleosides 5′-triphosphates The 5′-diphosphates and 5′-monophosphates cannot substitute for the 5′-triphosphates

The *E coli* DNA polymerase consists of about 1000 amino acid residues (MW 109,000) The enzyme appears to be an end-to-end protein, and there is no evidence for protein subunits or prosthetic groups There is evidence that within the active center of the enzyme there are several sites: *1*) a site for the template, *2*) a site for the primer, *3*) a site for the 3′-OH of the growing chain, and *4*) a site for the 5′-triphosphates (Fig 1 148). The synthesis proceeds from the 5′ → 3′ direction. A 5′-tri-

phosphate enters the appropriate site and is oriented so that it can be brought into direct contact and form a base pair with the template (Fig 1 148) When the correct base pair is formed, a nucleophilic attack by the 3′-OH on the innermost phosphate takes place and a



Fig 1 147 Autoradiography of an *E coli* chromosome *A* represents the moving fork Behind the moving fork one can observe the two separated daughter duplexes These duplexes are joined again at the initiation point *B* Reproduced from Cairns 1963

phosphodiester bond is formed At the same time a new 3'-OH replaces the old one and the enzyme moves (relative to DNA) and binds to the new primer terminus Again another 5'-triphosphate enters the triphosphate site and the polymerization step is repeated The reaction will be completed with the addition of the last base A (Fig 1 148) At that point it is evident that there is no template to continue polymerization in the 5' → 3' direction

DNA polymerase can quickly form a complete linear duplex from a DNA molecule as depicted in Figure 1 149*A* On the other hand, an intact linear duplex is inert and does not support replication since there is no template in the direction of 5' → 3' (Fig 1 149*B*) Similarly inert is a circular single or double stranded DNA

The Biological Function of DNA Polymerase Despite the accumulated knowledge on the structure and in vitro action of DNA polymerase, the involvement of this enzyme in the replication of DNA in vivo is still unclear However, DNA polymerase definitely plays a primary role in the DNA repair mechanism Several distinct types of DNA repair processes are known to exist The best known type is the repair of the thymine dimers (Fig 1 150) These dimers are formed by exposure of bacteriae cells to ultraviolet light If unrepaired, this damage is normally lethal It is proposed that repair proceeds with the following steps First, a specific endonuclease introduces a nick close to the damaged area, then an exonuclease excises a small region including the thymine dimer, the excised part is synthesized by DNA polymerase, finally, the last phosphodiester bond is catalyzed by the DNA ligase As one could expect, mutants which lack DNA polymerase are extremely sensitive to ultraviolet radiation Also, xeroderma pigmentosum, a human disease characterized by extreme sensitivity to ultraviolet light, involves defective DNA repair mechanisms This defect seems to be in the first step of the repair mechanism, namely the endonuclease step

Membrane Bound DNA-Replicase. Recent information suggests that the DNA-replicase is a cell membrane-bound enzyme Indeed isolated membrane fractions from bacteria catalyze DNA synthesis This enzymatic activity differs from that of DNA polymerase It is known that the bacterial chromosomes are attached to invaginated areas of the cell membrane, *the mesosomes* Isolated DNA cell membrane complexes from a DNA polymerase deficient mutant can catalyze the continuation of semiconservative replication of bacterial or phage DNA, and the rate of this reaction is comparable to rates of replication observed in vivo

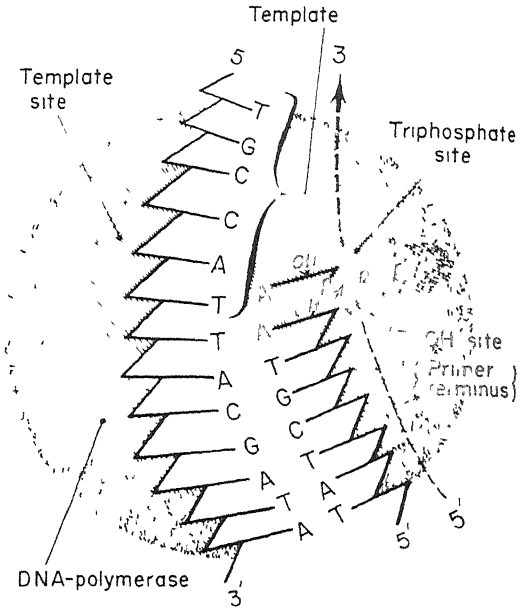


Fig 1 148 Binding sites and the mechanism of action of DNA polymerase A dATP molecule enters the triphosphate site and pairs with *T* This event triggers the nucleophilic attack of the 3' OH of the primer terminus on the innermost pyrophosphate bond of dATP A new phosphodiester bond will form and a new 3' OH primer terminus will appear A movement of the enzyme relative to DNA will bring the triphosphate site opposite to A of the template The polymerization step will be repeated with dTTP Note that the newly synthesized strand is antiparallel with the template strand Redrawn from Kornberg 1969

DNA Ligase. This enzyme can join the 3'-end with the 5'-end of two DNAs with the formation of a phosphodiester bond This enzyme can also join the two ends of a single stranded DNA to form a DNA circle The role of this enzyme in DNA replication is still uncertain However, the enzyme seems to participate in DNA repair processes as well as in the phenomenon of recombination of genes during crossing over

Replication of Single Stranded DNA Molecules. The replication of single stranded DNA molecules proceeds via the formation of a DNA duplex Phage $\phi \times 174$ for example possesses a single stranded circular DNA chro-

mosome, called the "+" *strand* Immediately after the "+" strand enters the host cell, it serves as a template for the synthesis of the complimentary "−" *strand* This synthesis is carried out by the enzymes of the host cell The phage DNA is now double stranded "+ − " The "+ −" duplex is called the *replicative form* The "−" strand of the "+ −" duplex serves as a template for the synthesis of "+" strands This phase of the life cycle of phage is carried out by a specific enzyme coded for by a phage gene The exclusive selection of the "−" strand as a template probably occurs because that strand contains specific base sequences which act as a signal and are recognized by the enzyme It is notable that recently Goulian, et al (1967) achieved the remarkable feat of enzymatically synthesizing a biologically active $\phi \times 174$ in a test tube

CONTROL OF DNA REPLICATION

DNA replication is an extremely complex process which requires and is under the control of a number of genes Thus, about 20 genes are involved for the replication of the relatively simple T4-phage Initiation of replication is dependent on protein biosynthesis It appears that certain protein(s) called initiation protein(s) act on a unique starting signal on DNA to begin a round of replication Thus, DNA itself, through specific genes, controls its own rate of replication

Structure, Synthesis, and Function of RNA

A given DNA molecule contains several genes According to accumulated genetic evidence, one gene codes for amino acid sequences of one polypeptide, and, therefore, we must assume that only one of the two DNA strands of the gene carries the genetic information Although DNA carries the genetic information, DNA itself is not the template on which proteins are synthesized Instead, DNA serves as a template for the synthesis of RNA by the enzyme, RNA *polymerase* This process is called *transcription* Only one strand of a gene is transcribed, that which carries the genetic information The RNA strand that is thus synthesized then serves as template for the synthesis of the protein (*translation*) The above statements can be summarized as follows

$$\text{DNA} \xrightarrow{\text{Transcription}} \text{RNA} \xrightarrow{\text{Translation}} \text{Proteins}$$

RNA STRUCTURE

The building blocks of RNA are the mononucleotides adenosine monophosphate (AMP), guanine monophosphate (GMP), cytosine

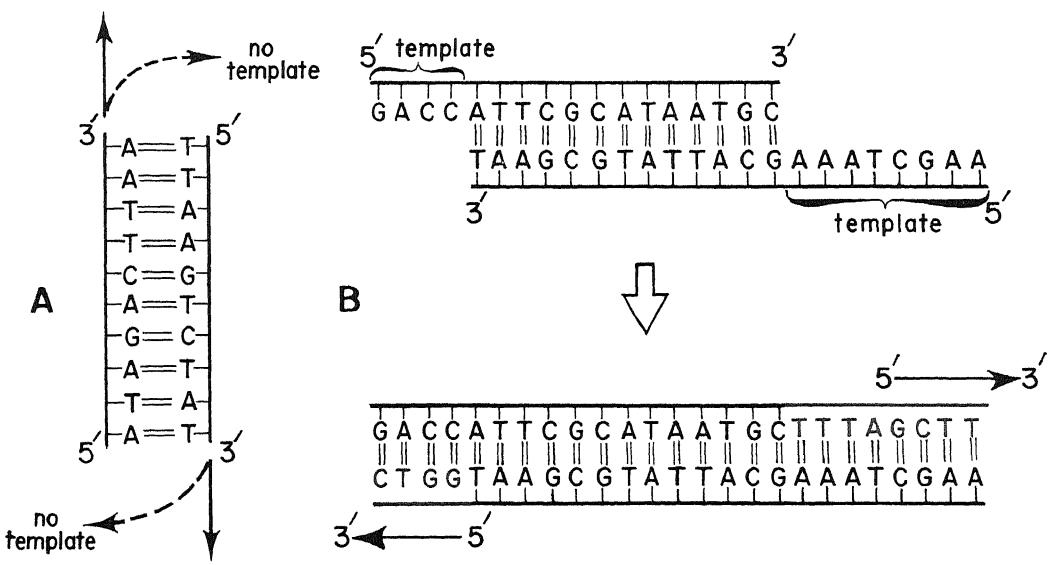


Fig 1 149 A complete linear duplex *A* is inert for replication by DNA polymerase because there is no template from which the enzyme can get instructions In contrast DNA molecules of the type depicted in *B* are excellent primer-templates for the polymerase

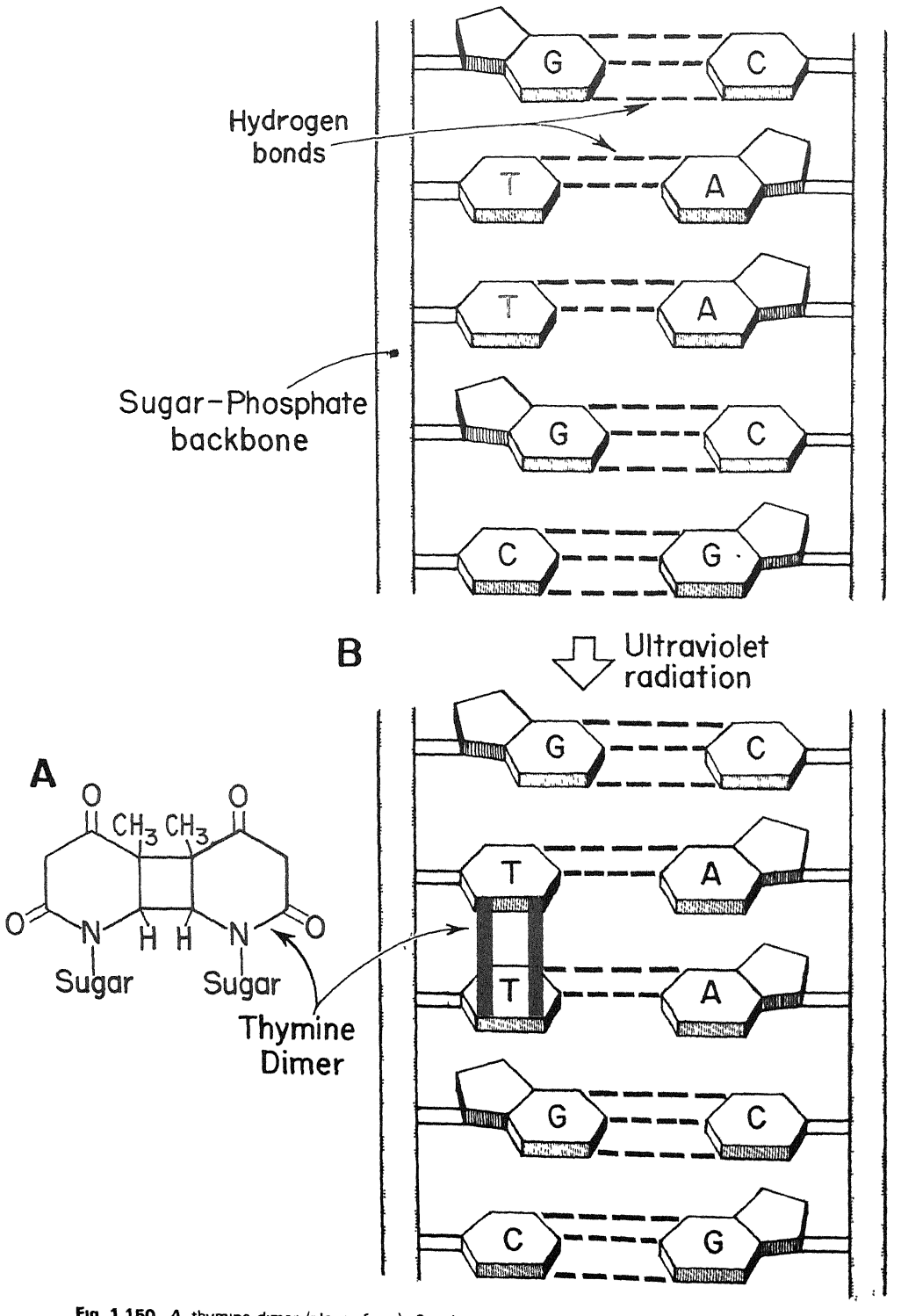


Fig 1 150 *A* thymine dimer (planar form) *B* schematic representation of thymine dimer formation

monophosphate (CMP), and uridine monophosphate (UMP) Note that in RNA uridine is found instead of thymine The fundamental units, the mononucleotides, are linked to one another with 3′,5′-phosphodiester bonds Thus in RNA as in DNA the covalent backbone is made of alternating ribose and phosphate groups To this backbone the purine and pyrimidine bases are attached as side-chains (Fig 1 151) RNA is usually single stranded In some cases RNA occurs in a form of two complementary strands which form a double helix In a duplex RNA the number of adenine residues is equal to the number of uridine residues, or A = U Similarly G = C and, therefore, A + G = C + U This composition can be explained if one assumes a Watson-Crick model for RNA in which the bases of one RNA strand are paired by hydrogen bonds to the bases of the other strand As in the case of DNA, only certain base pairs can be accommodated within the RNA duplex, these are A – U and G – C

RNA BIOSYNTHESIS

The biosynthesis of RNA is effected by the enzyme RNA *polymerase* For this synthesis,

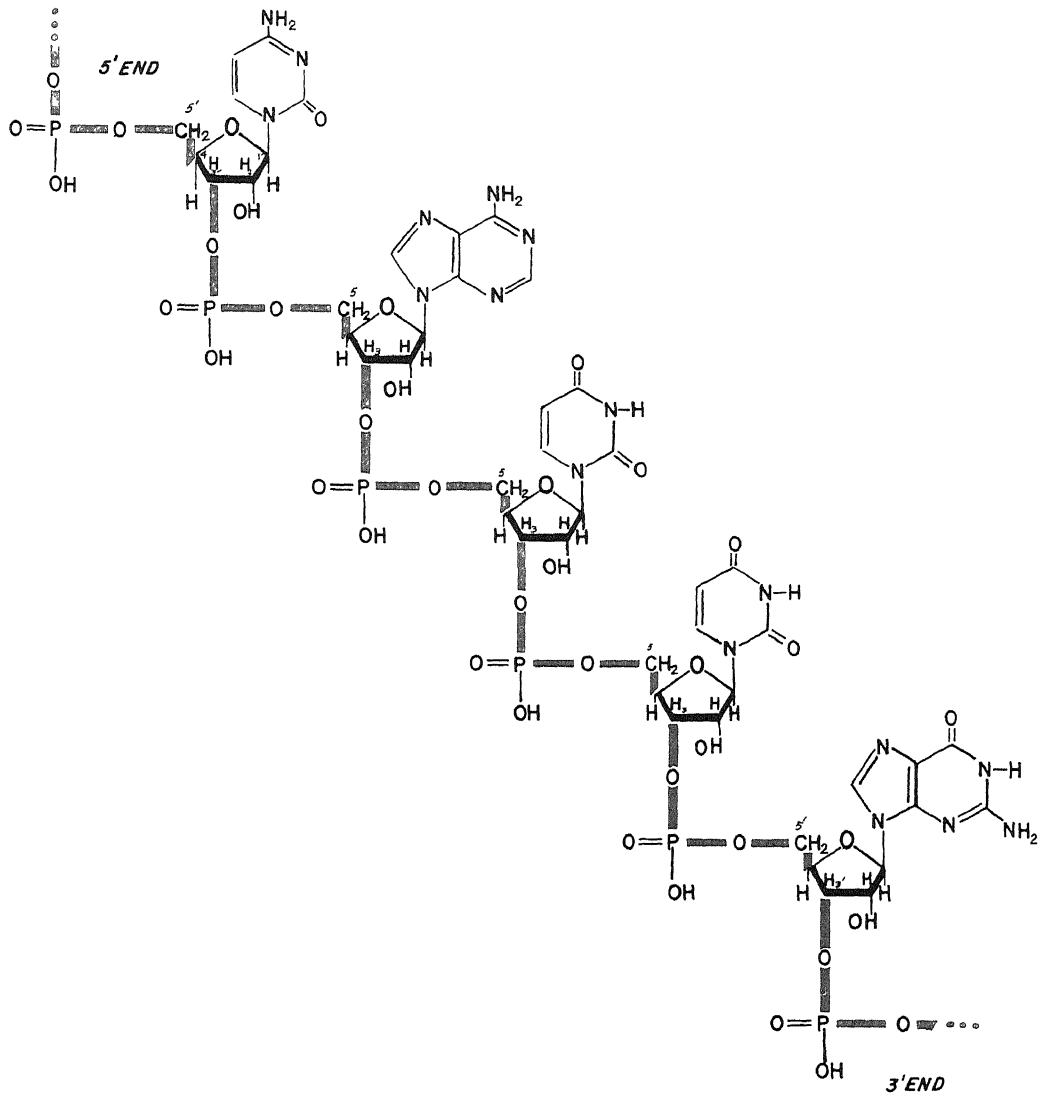


Fig 1 151 Structure of a portion of an RNA chain with hypothetical base sequence 5′ to 3′ cytosine-adenine uridine-uridine-guanine

the enzyme takes instructions from a DNA template The reaction catalyzed is the following

$$\begin{matrix} n_1ATP \\ n_2GTP \\ n_3CTP \\ n_4UTP \end{matrix} \xrightleftharpoons{\text{DNA } Mg^{++} \text{ Enzyme}} \begin{bmatrix} (n_1AMP) \\ (n_2GMP) \\ (n_3CMP) \\ (n_4UMP) \end{bmatrix} + (n_1 + n_2 + n_3 + n_4)PP$$

$$\downarrow \text{Pyrophosphatase}$$

$$2(n_1 + n_2 + n_3 + n_4)P_i$$

For each phosphodiester bond synthesized, a pyrophosphate bond is broken and a pyrophosphate group is released This reaction also has a slightly positive free energy The splitting of the high energy bond of pyrophosphate by the pyrophosphatase releases additional energy and ensures that the reaction is driven toward the direction of RNA synthesis The reaction requires all four ribonucleoside-5-triphosphates, as well as DNA to function as a template The 5'-diphosphates and 5'-monophosphates of ribonucleosides cannot substitute for the ribonucleoside 5'-triphosphates

Bacterial RNA polymerase is composed of protein subunits Five major subunits have been identified and designated as β', β, α, ω, and σ These subunits aggregate specifically with secondary bonds to form the complete enzyme or *holoenzyme* $\beta'\beta\alpha_2\omega\sigma$ (MW 470,000) Removal of σ (*sigma factor*) from the holoenzyme by column chromatography on phosphocellulose leads to the isolation of the $\beta'\beta\alpha_2\omega$ aggregate or *core enzyme* The core enzyme catalyzes the formation of the internucleotide phosphodiester bonds equally well as the holoenzyme Recent evidence suggests that the in vitro synthesis of RNA by the RNA polymerase, proceeds with the following steps *1) template binding, 2) chain initiation, 3) chain elongation,* and *4) chain termination* As noted above, only one strand of the DNA-duplex is transcribed in vivo, the strand that carries the genetic information Recent studies in vitro show that the holoenzyme also selects and transcribes only one strand In contrast, the core enzyme shows no strand selectivity with a variety of DNA templates Strand selection and binding of RNA polymerase to DNA template are effected through *start signals* (specific base sequences along the DNA molecule), such signals are recognized by the sigma factor

Once the RNA polymerase is bound to DNA template initiation of RNA synthesis appears to involve the binding of two 5' triphosphonucleotides to the enzyme This is followed by formation of a phosphodiester bond and elimination of pyrophosphate according to the reaction

$$ppp'A' + ppp'G' \rightarrow ppp\ ApG'$$

RNA chains start either with pppA or pppG when phage DNAs are used as templates Elongation of the initiated chain proceeds toward the direction 5' → 3' A nucleotide 5'-triphosphate enters the appropriate enzyme site and is oriented so that it can be brought into direct contact and form a base pair with the template (Fig 1 152) When the correct base pair is formed, a nucleophilic attack by the 3'-hydroxyl end of the RNA chain to the innermost phosphate takes place and a phosphodiester bond is formed Thus the polymerization reaction proceeds by the addition of 5'-triphosphates to the 3' hydroxyl end of the growing RNA chain The enzyme adds 40 to 100 nucleotide units per sec The pairing rules are ADNA pairs with URNA, TDNA pairs with ARNA, GDNA pairs with CRNA, and CDNA pairs with GRNA At some point of the elongation process in vitro the σ factor is released from the polymerase and the remaining core enzyme continues the polymerization This observation led to the proposal of the "*sigma cycle*" According to this postulate, the released σ factor combines with another core enzyme and initiates another round of transcription Thus, the σ factor may be critical in the regulation of the transcription process although, it is not known at the present time if the release of sigma factor also occurs in vivo (See section on regulation of gene activity below)

Once the strand selection is made, RNA polymerase moves quickly along the DNA molecule, separating the two strands over a short local region and utilizing one strand as a template The RNA as it is synthesized forms a transient hybrid duplex with its DNA template As the transcription of a region of the template is finishing, and the enzyme moves out of this region, the separated DNA strands come back together and form the original duplex with concomitant expulsion of the RNA from the hybrid duplex The expulsion of the RNA strand by the DNA strand occurs because

DNA-DNA duplexes are energetically more stable than DNA-RNA duplexes

Specific sequences of DNA (*stop signals*) induce the termination of the polymerization process Another termination mechanism involves the protein *factor ρ* The mechanism by which factor ρ triggers termination of RNA synthesis is not yet known

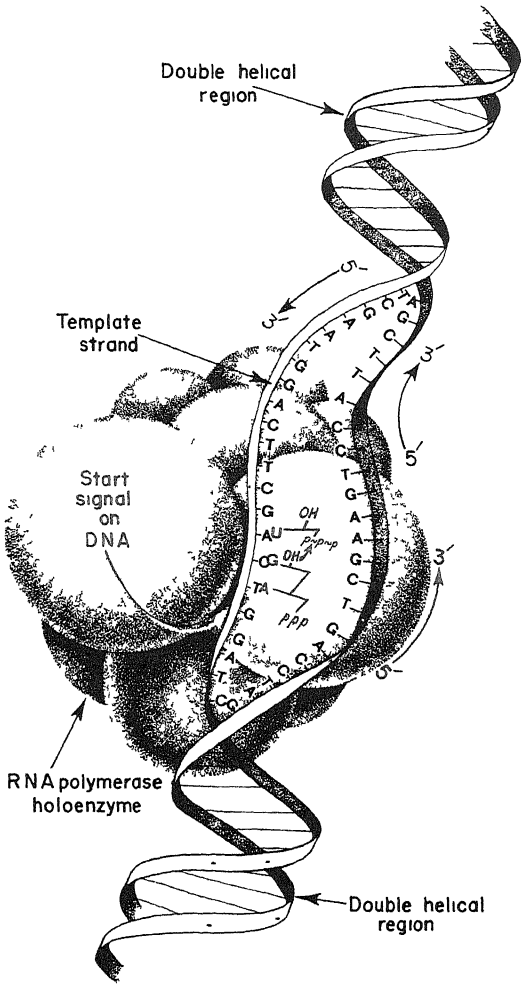


Fig 1 152 Schematic representation of DNA transcription by RNA polymerase The sigma factor of polymerase recognizes the start signal on DNA and the enzyme binds to template Binding at the start signal leads to opening of DNA strands over a short local region This is followed by binding on the enzyme of ATP and GTP and formation of the first phosphodiester bond (initiation) This yields the dinucleotide pppApG Initiation is followed by elongation A UTP molecule enters the appropriate site and pairs with *A* This event triggers the attack of 3′-OH on the innermost pyrophosphate bond of UTP A new phosphodiester bond will form and a new 3′-OH primer terminus will appear A movement of the enzyme relative to DNA will bring the triphosphate site opposite to *C* of the template The polymerization steps will be repeated again and again until a stop signal on DNA or a ρ-factor will induce the termination of the transcription process

Recent studies on RNA polymerases from eucaryotic cells reveal a variety of molecular forms of the enzyme These forms have been designated as RNA polymerase I, II, III, and IV RNA polymerase I is found in the nucleolus, II and III in the nucleoplasm, and IV in the mitochondria RNA polymerase I has a molecular weight in the range of 500,000 to 700,000, and this suggests that the enzyme most probably consists of subunits Similarly, RNA polymerase II (MW about 700,000) consists of several subunits Studies on the protein subunits of eucaryotic polymerases will probably reveal the existence of factors analogous to bacterial sigma factor through which template strand selection is effected

Inhibition of RNA Biosynthesis.

Actinomycin D, a peptide antibiotic, selectively inhibits DNA dependent RNA synthesis This inhibition is exerted in intact animals and cell cultures, as well as with purified enzyme systems, in vitro The observed inhibition is due to the binding of the antibiotic to DNA through hydrogen bonds This binding, because of stereochemical reasons, takes place between the phenoxazine ring of actinomycin and guanine residues in the minor groove of DNA As a result, the peptide chains of the antibiotic fill the minor groove of DNA and sterically prohibit the RNA polymerase from transcribing DNA

Rifamycin is another antibiotic which inhibits RNA biosynthesis Rifamycin binds specifically to the β-subunit of RNA polymerase and produces a block in RNA chain initiation The antibiotic *streptolydigin* inhibits RNA synthesis by blocking the chain elongation process Finally, the toxin α-amanitin reacts specifically with eucaryotic RNA polymerase II and blocks RNA synthesis

RNA Replicases (RNA-dependent RNA polymerases) RNA replicases are the enzymes of replication of the RNA viruses The enzymes are coded by viral genes and catalyze the formation of RNA strands from the ribonucleoside-5′-triphosphates on RNA templates, thus they are referred to as the name RNA dependent-RNA polymerases

The replicases are usually formed immedi-

ately after the infection of a cell by an RNA virus. A replicase is highly specific for the RNA virus by which it is coded. Only this viral RNA serves as a template for the enzyme, the specificity is due to specific nucleotide sequences of the RNA which are recognized by the enzyme.

Recent studies have uncovered yet another enzyme, the *RNA-dependent* DNA polymerase or reverse transcriptase. This enzyme catalyzes the synthesis of DNA on RNA templates. The precise role of this enzyme is not yet known.

THE VARIOUS RNA CLASSES AND THEIR FUNCTION

Messenger RNA (mRNA) In 1961 Jacob and Monod introduced the messenger RNA hypothesis. They proposed that genetic messages of DNA are transcribed faithfully into messenger RNA whose function is to carry the information to the site of protein synthesis and serve as a direct template for that synthesis. Since then the existence of mRNA has been verified experimentally. The mRNAs are, therefore, the transcription products of genes. An mRNA can contain the complete message of one gene coding for one polypeptide chain or of several genes coding for several polypeptide chains. In the latter case, the mRNA is called *polygenic* or *polycistronic*. The mRNAs represent an heterogeneous population of RNA molecules of different chain length. This heterogeneity is expected since the various mRNAs are the templates of different proteins. To separate and identify a specific messenger RNA is an extremely difficult task because of the heterogeneity of mRNAs and because mRNAs constitute less than 1% of the total cellular RNA. However, certain messengers such as the mRNAs coding for chains of hemoglobin and various viral mRNAs have been successfully isolated.

Ribosomes and Ribosomal RNA. Cell fractionation of liver tissue obtained from rats injected with radioactive amino acids enabled Zamechnik (1960) to identify the intracellular site of protein synthesis. It was found that protein synthesis is associated with the cellular fraction of *microsomes* (Fig. 1-153). Moreover, incubation of microsomes in vitro, in the presence of the cytoplasmic *soluble fraction* (cytosol), ATP, GTP, and labeled amino acids, results in amino acid incorporation into proteins. This in vitro system constitutes a *cell free system* for protein synthesis. Treatment of microsomes with sodium deoxycholate yields ribonucleoprotein particles, the *ribosomes*, which can be isolated by centrifugation. Most of the ribosomes in the eucaryotic cell are attached to the endoplasmic reticulum, although some are free in the cytoplasm (Fig. 1-154). Ribosomal preparations can substitute for microsomes in the cell free system of protein synthesis. Actually, ribosomes by themselves cannot support protein synthesis, indeed the active units of protein synthesis in ribosomal preparations are aggregates of ribosomes called *polysomes*. Polysomes are formed from the specific association of a mRNA molecule with ribosomes.

A ribosome consists of two subunits, one is large and one small. Bacterial ribosomes have a sedimentation value of 70S (Svedberg units) and their subunits have values of 50S and 30S. Eukaryotic ribosomes have a sedimentation constant of 80S and their subunits have values of 60S and 40S. Ribosomes, suspended in buffers containing less than 10^{-4} M Mg^{++}, dissociate into their subunits, raising the Mg^{++} concentration to 5×10^{-4} M causes the subunits to associate again into functional ribosomes. Dissociation and association of ribosomes and their subunits also occurs in vivo in association with the process of protein synthesis, and involves special factors (see chap. 5). The

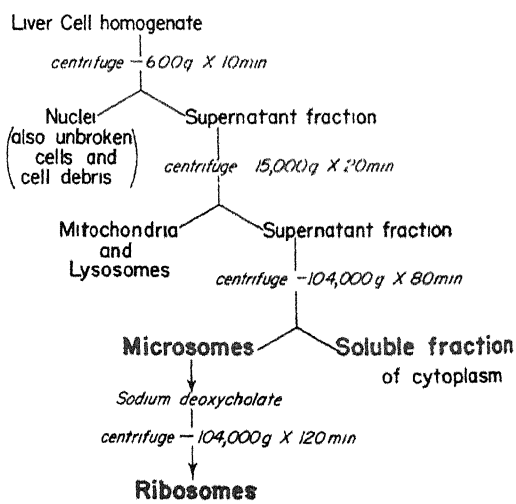


Fig. 1-153 Fractionation of liver cells by differential centrifugation. Cells are homogenized in isotonic sucrose buffers pH 7.4 containing Na^+, K^+, and Mg^{++}. For the isolation of "active" microsomes or ribosomes all operations must be performed at 0.2°.

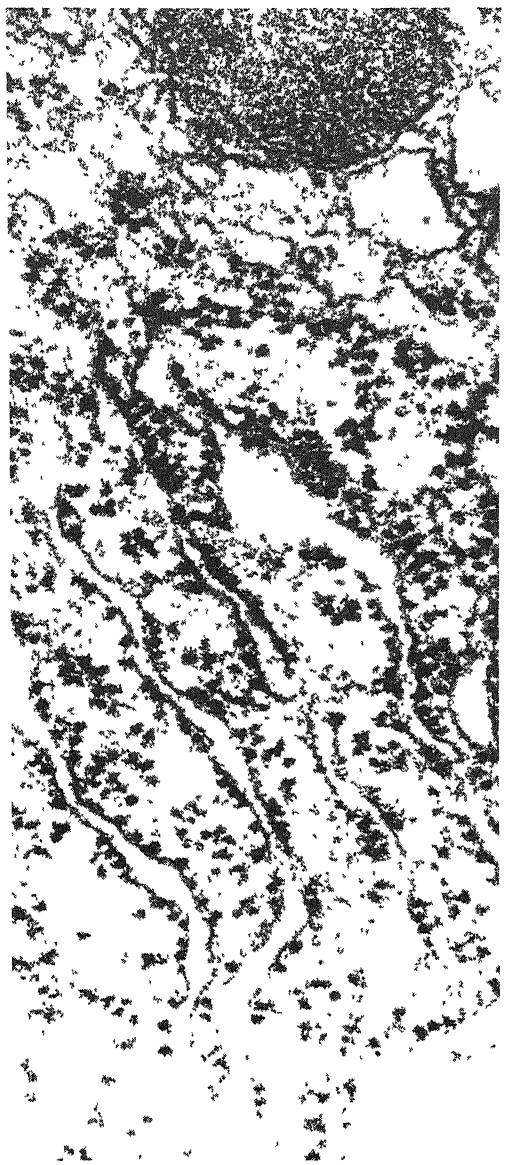


Fig 1 154 Electron micrograph of a section of rat liver cell Most of the ribosomes are attached to the endoplasmic reticulum and some are free in the cytoplasm

ribosomal subunits are composed of ribosomal RNA (rRNA) and proteins The 50S subunit consists of one molecule of 23S rRNA (MW 1,200,000), one molecule of 5S rRNA and about 35 different proteins The 30S subunit, consists of one molecule of 16S rRNA (MW 600,000) and 20 different proteins Under certain conditions, mixing of the separated proteins from the 30S subunit together with 16S rRNA leads to reconstitution of biologically active 30S subunits (Nomura, 1973) This demonstrates that the complex 30S subunit structure (and most probably the ribosome itself) represents a self assembling system, in which the various components bind to one another in some specific manner and assume the proper conformation The rRNAs from subunits of eukaryotic ribosomes have sedimentation values of 28S, 18S, and 5S

The ribosomal RNAs are synthesized on DNA templates A *ribosomal gene* codes both the 23S (28S) and 16S (18S) rRNAs In a given cell there are several copies of the ribosomal genes In fact the number of ribosomal genes increases from bacteria to insects, to mammals, to amphibians, this increase roughly parallels the observed increase of DNA per cell from species to species For example, mycoplasma possesses a single ribosomal gene, and bacteria have from 5 to 10 In eukaryotic cells the number increases tremendously reaching tens of thousands in some amphibians which possess the highest amount of DNA per cell The ribosomal genes of eukaryotes are serially repeated and clustered in one locus in the nucleolar DNA Hybridization experiments show extensive cross reaction of bacterial ribosomal RNA with DNAs from eukaryotic cells This constitutes evidence that ribosomal genes were conserved to a large extent over the course of evolution

The large number of ribosomal genes is most probably necessary to provide sufficient template to satisfy the demand for ribosomes In eukaryotic cells, ribosomal genes are transcribed to give a *precursor RNA* (45S) which undergoes specific cleavages to yield both the 28S and the 18S rRNA

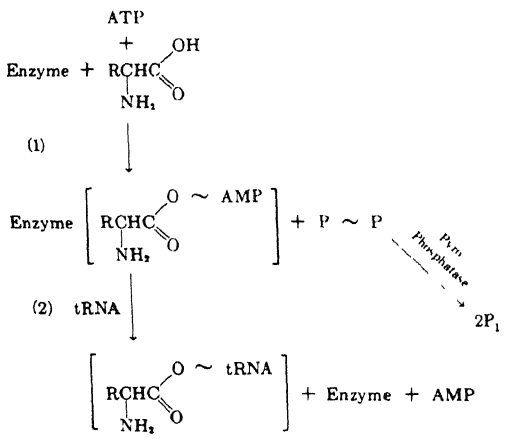
Transfer RNA (tRNA) and the Activation of Amino Acids There are many tRNAs in a cell These tRNAs bind covalently with the various amino acids, and then participate in the process of protein synthesis A given tRNA is specific for a given amino acid The tRNAs have a MW of about 25,000 Each tRNA is composed of a single chain of approximately 80 nucleotides, covalently linked, with many hydrogen bonded base pairs This base pairing causes the chain to form several loops

Elucidation of the primary structure of yeast alanine tRNA was achieved in 1965 by Holley and his colleagues Since then other tRNAs from various species have also been sequenced When tRNA sequences are written in a way

that produces the highest number of base pairs, a 'cloverleaf" form appears (Fig 1 155) This cloverlead form consists of four 'loops", one of which contains the *anticodon* site (see below "Genetic Code") All tRNAs terminate at the 3'-end with the —CCA sequence This is the site of amino acid attachment

The tRNAs contain in their structure many unusual bases, the precise role of which is not yet known The three-dimensional structure of tRNA has not yet elucidated However, now that tRNAs have been obtained in crystalline form, the elucidation of their 3-D conformation through x-ray crystallography appears feasible The various tRNAs are coded by specific genes and, therefore, are synthesized on DNA templates

Since the naturally occurring amino acids in proteins are 20, there must be at least 20 kinds of tRNA molecules Actually there are more than 20 (see section on degeneracy of the genetic code) A transfer RNA binds the appropriate amino acid with a single covalent bond This reaction is catalyzed by specific amino acyl synthetases as follows



The reaction proceeds in two steps In the first step ATP and the amino acid react on the catalytic site of the enzyme to give the enzyme bound intermediate amino acid ~ AMP During this reaction a high energy pyrophosphate bond is broken and a high energy anhydride

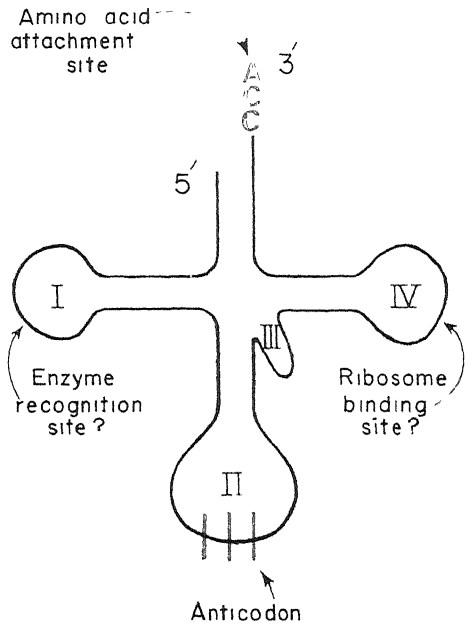


Fig 1 155 Standard orientation of a tRNA in the cloverleaf form

linkage is formed In the second step the amino acyl group is transferred from AMP to the 3'-OH end of tRNA The ester linkage between the amino acid and tRNA is a high energy bond The splitting of the high energy bond of pyrophosphate by the pyrophosphatase releases additional energy This event ensures that the amino acid activation reaction is driven to completion The presence of Mg^{++} in the reaction is an absolute requirement Each amino acid is activated by a specific amino acyl synthetase Almost all the synthetases have been isolated and all possess one or more —SH group which is essential for catalytic activity All synthetases have a molecular weight of about 100,000 with the exception of phenylalanine amino acyl synthetase which has about 180,000

The synthetases must possess at least three binding sites, *a*) specific site for the amino acid, *b*) specific site for tRNA, and *c*) site for ATP Obviously, each tRNA must also possess an *enzyme recognition site* This site has not yet been identified and located in the cloverleaf form of sequenced tRNAs, although it is suspected to be in the area of loop I (Fig 1 155) The very high specificity of the enzyme for a particular amino acid and its appropriate tRNA makes it very unlikely in vivo ($p <$

0 0001) that a tRNA could be charged with the "wrong" amino acid

The Genetic Code

In as much as the genetic information for ordering amino acids in polypeptide chains, is stored in DNA in the form of base sequences, a relationship between bases and amino acids must exist Since there are only four bases and twenty amino acids, a one-to-one or even two-to-one correspondence between bases and amino acids is not possible A three-to-one correspondence on the other hand yields 64 base triplets to code for 20 amino acids This constitutes a *triplet code* in which the three consecutive bases coding for one amino acid constitute a *codon* Crick and his colleagues obtained evidence for such a triplet code from genetic experiments Pioneering work by Nirenberg and his colleagues resulted in the complete elucidation of the code This feat was achieved experimentally by the use of artificial mRNAs, which permitted the complete assignment of codons to the various amino acids

ENZYMATIC SYNTHESIS OF POLYNUCLEOTIDES (ARTIFICIAL MESSENGERS)

Synthetic polynucleotides can be obtained enzymatically with the use of the enzyme *polynucleotide phosphorylase* This enzyme, found in bacteria, catalyzes the following reaction

$$\begin{matrix} n_1\text{ADP} \\ n_2\text{GDP} \\ n_3\text{CDP} \\ n_4\text{UDP} \end{matrix} \xrightarrow[\text{Mg}^{++}\ \text{Enzyme}]{\text{RNA primer}} \text{RNA}-\begin{bmatrix} (\text{AMP})_{n_1} \\ (\text{GMP})_{n_2} \\ (\text{CMP})_{n_3} \\ (\text{UMP})_{n_4} \end{bmatrix} + (n_1 + n_2 + n_3 + n_4)\text{P}_i$$

The reaction requires the 5′-diphosphates of ribonucleosides, the enzyme cannot utilize the 5′-diphosphates of deoxynucleosides In the reaction, a primer RNA strand furnishes the 3′-hydroxyl end of which the enzyme adds additional residues However, because the enzyme does not utilize a template, the residues added do not follow a specific sequence pattern Using only uridine 5′-diphosphate in the reaction, one obtains a polyuridylic acid product (*poly U*) By analogous means, one can prepare *poly A*, *poly C*, and *poly G* By using two of the diphosphonucleotides, for example 5′-UDP and 5′-GDP, one can prepare mixed copolymers *poly U, G* The base composition of the copolymers reflects the ratio of the original concentrations of the UDP and GDP precursors Thus, if the original ratio was 5 1 UDP to GDP then the poly U, G copolymers contain five times more U residues than G

ASSIGNMENT OF CODONS

In 1961, Nirenberg and Matthaei discovered that incubation of poly U together with ribosomes and the other factors of the cell free system for protein synthesis promotes the incorporation of phenylalanine into protein Phenylalanine was the only amino acid incorporated out of 19 tested Further investigation revealed that the product formed was polyphenylalanine Thus, it became apparent that poly U codes for phenylalanine and the codon UUU was tentatively assigned to this amino acid From subsequent experiments with poly A and poly C, the codon AAA was assigned to lysine and the codon CCC to proline This line of experimentation was pursued with mixed copolymers such as poly U, G Poly U, G contains the following eight codons UUU, UUG, UGU, GUU, UGG, GUG, GGU, and GGG The relative frequencies of occurrence of these codons can be calculated because it depends on the U G ratio Thus, in a poly U, G (5 1) the relative frequencies of occurrence are UUU = 100, (UUG, UGU, GUU) = 20, (UGG, GUG, GGU) = 4, and GGG = 0 8 Poly U, G (5 1) as expected promotes the incorporation of several amino acids in a cell free system The amino acids and their relative amount of incorporation were found to be phenylalanine = 100, valine = 20, cysteine = 20, leucine = 20, glycine = 4, and tryptophan = 4 It appears then that the codon for phenylalanine is UUU, the codon for valine, cysteine, and leucine is U_2G and the codon for glycine and tryptophan is UG_2

Such experiments with mixed copolymers permitted the determination of composition of the codon of several amino acids but did not allow the deduction of the sequences of the different nucleotides within the assigned codons Thus, there is not way to tell whether the codon for valine which has a composition of U_2G is UUG, UGU, or GUU Determination of the nucleotide sequences of codons was

achieved with the aid of the binding assay of Nirenberg and Leder (see Nirenberg, 1968) The assay rests on the specific binding of an amino acyl-tRNA on the mRNA-ribosome complex Thus poly U mixed with ribosomes binds only phenylalanyl-tRNA More important was the finding that instead of poly U, one can use the trinucleotide UpUpUp and still observe specific binding The use of the assay with trinucleotides of known sequences resulted in the determination of the nucleotide order in the codons Thus, it was found that the trinucleotide GpUpUp promotes valyl-tRNA binding, establishing that the codon for valine with the base composition U_2G is actually GUU

Development of techniques for the synthesis of polyribonucleotides of known sequences provided another tool for the study of the coding problem (Khorana, 1968) For example, the use of the synthetic messenger CpUpC p UpCpU p CpUpC containing the alternating codons CUC and UCU was found to code for a polypeptide in which serine and leucine residues alternate Thus, with the use of homopolymers, mixed copolymers, copolymers of known sequences, and trinucleotides the genetic code was totally deciphered (Table 1 13)

The genetic code is degenerate, i e , there is more than one codon assigned to one amino acid Thus, UUU and UUC are both assigned to phenylalanine and the codons GUU GUC, GUA, and GUG are assigned to valine Degeneracy in the case of the phenylalanine and valine codons involves the third base This type of degeneracy is the most frequent Some amino acids have codons with identical bases in the first two positions In all these cases, the third position is filled *only* by purines or *only* by pyrimidines For example, asparagine is coded for by AAU and AAC, and lysine by AAA and AAG Note that in the case of asparagine the third position is filled only by the pyrimidines, U or C, whereas in the case of lysine the third position is filled only by the purines, A or G

Sixty-one out of 64 codons have been assigned to the twenty amino acids while the remaining three UAA (ochre), UAG (amber), and UGA (umber) do not correspond to any amino acid (*nonsense codons*) The nonsense codons are signals for polypeptide chain

Table 1 13
Codon assignments for amino acids

	U	C	A	G
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys
	UUA Leu	UCA Ser	UAA Ochro	UGA (Umber)
	UUG Leu	UCG Ser	UAG Amber	UGG Trp
C	CUU Leu	CCU Pro	CAU His	CGU Arg
	CUC Leu	CCC Pro	CAC His	CGC Arg
	CUA Leu	CCA Pro	CAA Gln	CGA Arg
	CUG Leu	CCG Pro	CAG Gln	CGG Arg
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser
	AUC Ile	ACC Thr	AAC Asn	AGC Ser
	AUA Ile	ACA Thr	AAA Lys	AGA Arg
	AUG Met	ACG Thr	AAG Lys	AGG Arg
G	GUU Val	GCU Ala	GAU Asp	GGU Gly
	GUG Val	GCC Ala	GAC Asp	GGC Gly
	GUA Val	GCA Ala	GAA Asp	GGA Gly
	GUG Val	GCG Ala	GAG Glu	GGG Gly

termination From all evidence to date, it appears that the genetic code is *universal* for all organisms

RECOGNITION OF THE CODONS THE ANTICODON

It is now well established that the insertion of amino acids into the growing polypeptide chain is mediated through tRNA molecules Furthermore, codon recognition is not effected by the amino acid, but it rests with the tRNA This was proved elegantly with the following experiment Cysteine was charged enzymatically to the appropriate tRNA

$$\text{Cysteine} + \text{ATP} + \text{tRNA}_{\text{cyst}} \xrightleftharpoons[\text{Amino acyl Synthetase Mg}^{++}]{} \text{Cysteinyl} - \text{tRNA}_{\text{cyst}} + \text{AMP} + \text{PP}$$

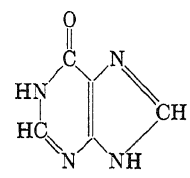
The cysteinyl – $tRNA_{cyst}$ was then transformed by reduction, to alanyl – $tRNA_{cyst}$ Thus, an activated alanine was produced in attachment to the tRNA for cysteine Cell free systems from reticulocytes incorporate amino acids into the hemoglobin chains Use of alanyl -$tRNA_{cyst}$ in such a system afforded the isolation of hemoglobin chains having alanine in the positions regularly occupied by cysteine The implication of this experiment is that the specificity of codon recognition of the amino acyl tRNA is in the tRNA, and if a tRNA molecule is charged with the "wrong" amino acid, this will produce translational "errors " Essentially, these kinds of error do not occur in vivo because there is a high degree of specificity of the amino acid activation reaction

Recognition of the codons is effected by the anticodons on the tRNA molecules, these anticodons are composed of three nucleotides The recognition was originally thought to be through regular Watson-Crick base pairs formed by the interaction of a codon with its anticodon This concept, which requires a specific anticodon for each of the sixty-one codons and, therefore, a specific tRNA for each codon, was not substantiated experimentally for it was found that a tRNA could recognize more than one codon To explain these observations, the "wobble hypothesis" has been introduced According to this hypothesis, the codon-anticodon interaction involves hydrogen bond formation due to base pairing Starting from the 3'- of the anticodon, the first two nucleotides form regular Watson-Crick type base pairs with the two nucleotides of the codon starting from the 5'- The third nucleotide at the 5'- end of the anticodon has more freedom of movement and thus pairing is not necessarily confined to the known pairing rules (Table 1 14) Thus, the two codons of tyrosine, UAU and UAC, are both recognized by one tRNA bearing the anticodon $^{3'}AUG^{5'}$ The "wobble" at the third position permits G of the anticodon to pair with either C or U (Fig 1 156)

Table 1 14
Pairing rules at the wobble position (Crick 1966)

Nucleotide at the "Wobble" Position of the Anticodon (5'-end)	Nucleotide at the Third Position of the Codon (3'-end)
Adenine	Uracil
Guanine	Uracil or Cytosine
Uracil	Adenine or Guanine
Cytosine	Guanine
Inosine	Adenine, Uracil, or Cytosine

In many cases the third base of the anticodon is found to be *Inosine* (I)



Inosine

The "wobble" hypothesis predicts that I could pair with A, U, or C Thus, three of the four codons of valine, GUA, GUU, and GUC are recognized by one tRNA bearing the anticodon 'CAI (Fig 1 157)

Support in favor of the "wobble" concept has steadily accumulated with the elucidation of sequences of more and more tRNAs and identification of their anticodons Up to now more than 12 anticodons have been identified, none of which contradicts the "wobble" concept

GENETIC EVIDENCE FOR THE CODE

Single amino acid changes in proteins result from a mutation involving change of one nu-

cleotide within one codon Consider for example the underlined single base changes in the following codons

1 (Leu) CUU → UUU (Phe)
2 (Ser) UCU → UUU (Phe)
3 (Tyr) UAC → UUC (Phe)

Replacement of leucine by the phenylalanine was found to occur at position 10 of mutant-TMV protein Replacement of serine by the phenylalanine was found to occur at position 138 and 148 of mutant-TMV protein Finally replacement of tyrosine by phenylalanine was found in position 35 of the β-chain of an abnormal hemoglobin From the catalogue of the assigned 64 codons (Table 1 12) it can be calculated that the possible single base changes that can lead to another amino acid are 75 Up to now, 54 of these possible changes have been identified in amino acid replacement data from various proteins

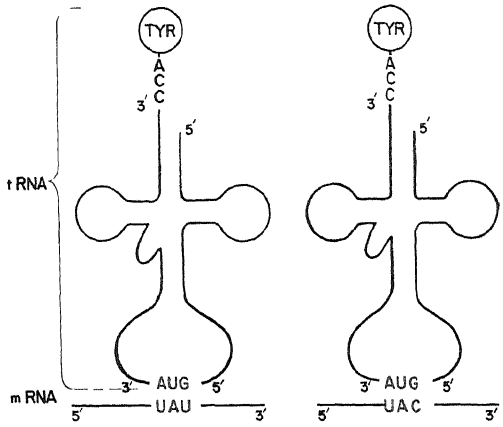


Fig 1 156 Pairing of codon and anticodon in a diagram that illustrates significance of wobble at the third position

Confirmation of the Code from Phage R17 RNA Sequences Advances in the biochemical technology of nucleotide sequence determination provided the tool for the dramatic elucidation of sequences of a viral RNA Sequence studies on the phages R17 and Qβ are now in progress

The nucleotide sequence of the gene for the coat protein of R17 phage has been already elucidated and affords direct comparison with the known amino acid sequence of that protein The findings are consistent with the concept of a triplet-degenerate code In the sequence of the coat protein gene, 38 different codons were encountered, they code for amino acids in agreement with the previously described assignments

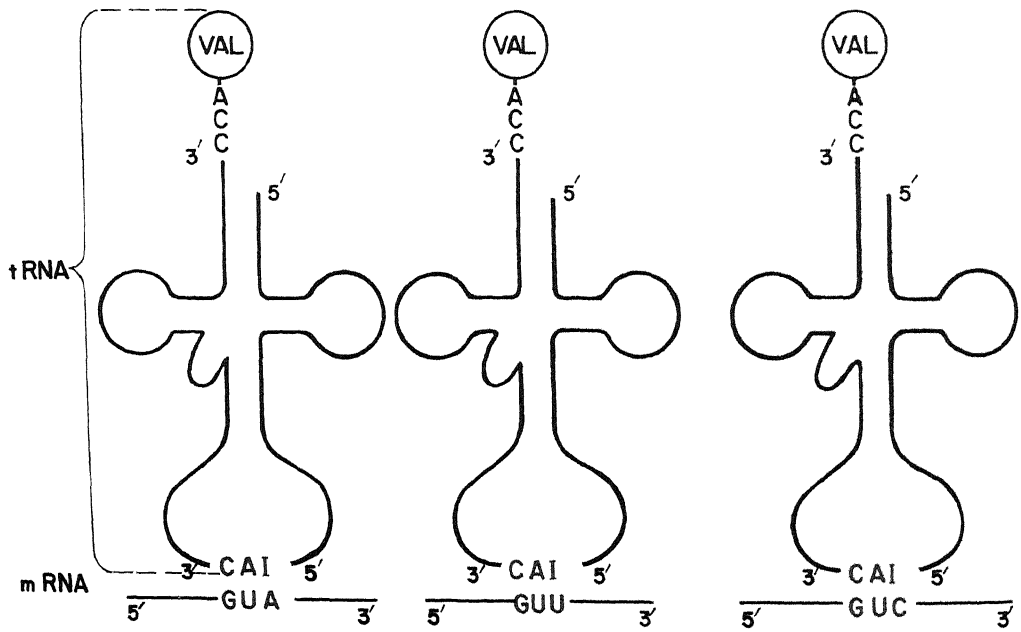


Fig 1 157 Codon-anticodon pairing that illustrates three possible combinations at the third position based upon wobble

Function of RNA in Protein Biosynthesis

The biosynthesis of proteins on mRNA templates is an elaborate process during which the genetic message is faithfully translated into a polypeptide chain The direction of translation of mRNA is from 5′ → 3′ The synthesis of a polypeptide chain starts from the amino terminal end and proceeds toward the carboxyl terminal end Peptide biosynthesis appears to proceed in a series of steps These have been designated as *1*) peptide chain *initiation*, *2*) peptide chain *elongation*, and *3*) peptide chain *termination*

PEPTIDE CHAIN INITIATION

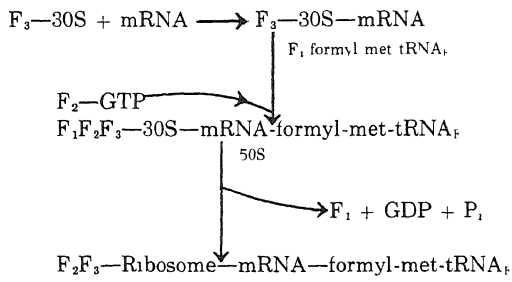
The process of peptide chain initiation in bacteria involves formyl-methionyl-$tRNA_F$ (F-met-$tRNA_f$), ribosomal subunits, the initiation signal of mRNA, the initiation factors F_1, F_2, F_3, and GTP The majority, if not all, bacterial polypeptides are initiated with *N-formyl-methionyl residues* The formylation of methionine takes place after the formation of methionyl-tRNA There are two tRNAs for methionine, $tRNA_F$ and $tRNA_M$ Methionyl-$tRNA_f$ can be formylated by transfer of a formyl residue from N^{10}-formyl-tetrahydrofolate to methionyl-$tRNA_F$

Transformylase
N^{10}-formyl-tetrahydrofolate

Methionyl-$tRNA_M$ cannot be formylated enzymatically and its function is to insert methionine residues at internal positions in the polypeptide chain

Although peptide chain initiation starts with formyl-methionine, bacterial proteins isolated from the cells do not have this residue at the *N*-terminus and frequently do not even start with methionine This occurs because the formyl residue is removed by the specific enzyme, deformylase, after the peptide chain has been initiated Methionyl residues are cleaved by another specific enzyme not yet isolated

Both the $tRNA_F$ and $tRNA_M$ have been sequenced recently and both bear the same anticodon [3] UAC[5] This anticodon matches the methionine codon AUG which is the *initiation* codon Initiation begins with the formation of initiation complex The complex is formed by the 30S and 50S ribosome subunits, the mRNA, the F-met-$tRNA_F$, three protein *initiation factors* F_1, F_2, F_3, and GTP as follows



Factor F_3 is required for the binding of 30s subunit to mRNA, whereas factor F_1 promotes the binding of F-met-$tRNA_f$ to the initiation complex Factor F_2 catalyzes the cleavage of GTP to GDP and inorganic P in the presence of ribosomes

Although met-$tRNA_M$ possesses the same anticodon as F-met-$tRNA_F$, it does not bind to the initiation codon AUG, it binds to the AUG codon only when this codon specifies methionine residues within the polypeptide chain It appears then that the specificity for the binding at the initiation position lies with the initiation factors, especially F_1 which interacts only with the F-met-$tRNA_f$ and not with the met-$tRNA_M$ Conversely, met-$tRNA_F$ (without the formyl group) does not participate in the peptide elongation process because it is discriminated by the T_u factor (see "elongation process")

After initiation, the direction of translation of mRNA is from 5′ → 3′ This does not mean that the initiation codon must be at the 5′- end Thus, the mRNA may possess, starting from the 5′- end, a number of nucleotides which precede the initiation codons Those immediately preceding the initiation codon do not have any coding function but have an initiation function Indeed, this area may represent a "*sticky point*" for the attachment of the 30S subunit as shown in Figure 1 158 In a polygenic messenger there may be the picture showing in Figure 1 159 A 30S subunit may attach directly to "*sticky point*" next to protein B and initiate the translation of protein B, thus by passing protein A

ELONGATION OF POLYPEPTIDE CHAINS

At the completion of the initiation process a ribosome is properly registered with the mRNA, and the initiator F-met-$tRNA_f$ is in

the *peptidyl site* ("P") of the complex (Fig 1 160*A*) The "P" site is formed by participation of the 30S and 50S subunits as well as the initiation codon At that state of the initiation the codon next to the initiation codon is a part of the amino acyl site ("A") of the ribosomes Site "A," as site "P," is formed by participation of the 30S and 50S subunits and the mRNA codon which is next in line to be decoded (Fig 1 160*A*) The subsequent events which ultimately will lead to peptide bond synthesis take place in three steps

Step One Binding of the Amino Acyl-tRNA In that step, the proper amino acyl-tRNA enters and binds to the "A" site (Fig 1 160*B*) This binding is effected by a protein factor The factor isolated from *E coli* is designated as T_u factor The T_u factor forms a complex with amino acyl-tRNA and GTP T_u + amino acyl-tRNA + GTP → T_u-amino acyl-tRNA-GTP The appropriate T_u complex binds to the "A" site to match the codon and immediately GTP breaks down to GDP and P_i which are released together with the binding enzyme (Fig 1 160*B*) The same T_u factor binds apparently all amino acyl-tRNAs The T_u factor does not bind to uncharged tRNAs, F-met-$tRNA_i$ and met-$tRNA_i$ (deformylated) Factors analogous to T_u exist also in eukaryotic systems, for example the *amino acyl transferase* I which has been isolated from liver tissue

Step Two· Peptide Bond Formation In this step a peptide bond will be formed between the amino group of the incoming amino acid and the activated carboxyl terminal of F-methionine (Fig 1 160*B* and *C*) This reaction is catalyzed by the enzyme *peptidyl transferase* which is an integral part of the 50S subunit

It appears that the enzyme has at least two binding sites, *a*) a site for all -CCA termini of the peptidyl-tRNAs, which is also the site of the initiator-tRNA, and *b*) a second site for the -CCA termini of the aa-tRNAs The first site is always filled by peptidyl tRNAs except at initiation of protein synthesis when it is filled by F-met-$tRNA_i$ Note also that in both cases (peptidyl or initiator tRNA) the α amino group of the amino acid attached to the tRNA is not free Therefore, one of the basic differences in specificity between the two binding sites of the

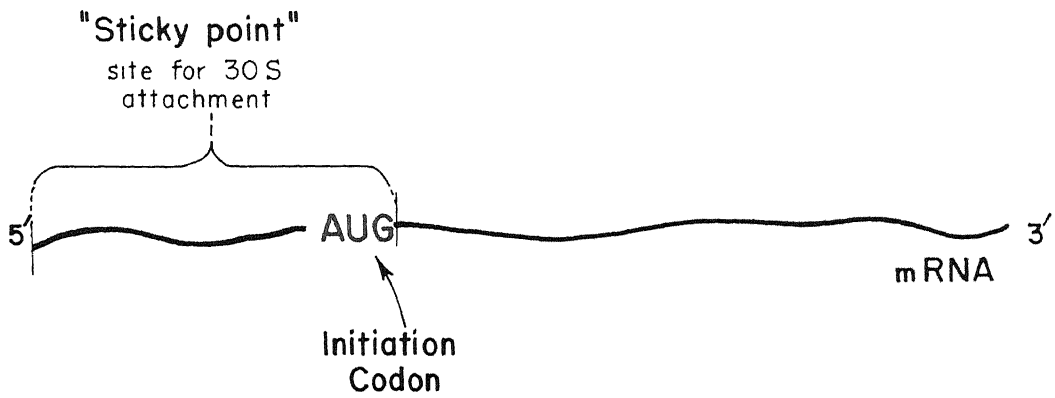


Fig 1 158 Diagram to illustrate location of 'sticky point' for attachment of a 30 S subunit in region of nucleotides possessing an initiation function

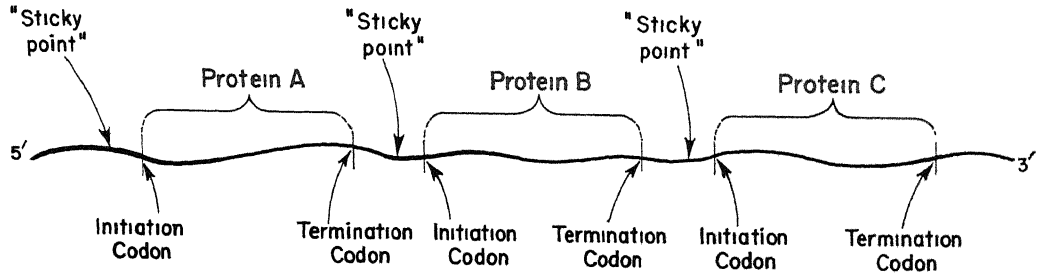


Fig 1 159 Location of 'sticky points' in a polygenic messenger RNA

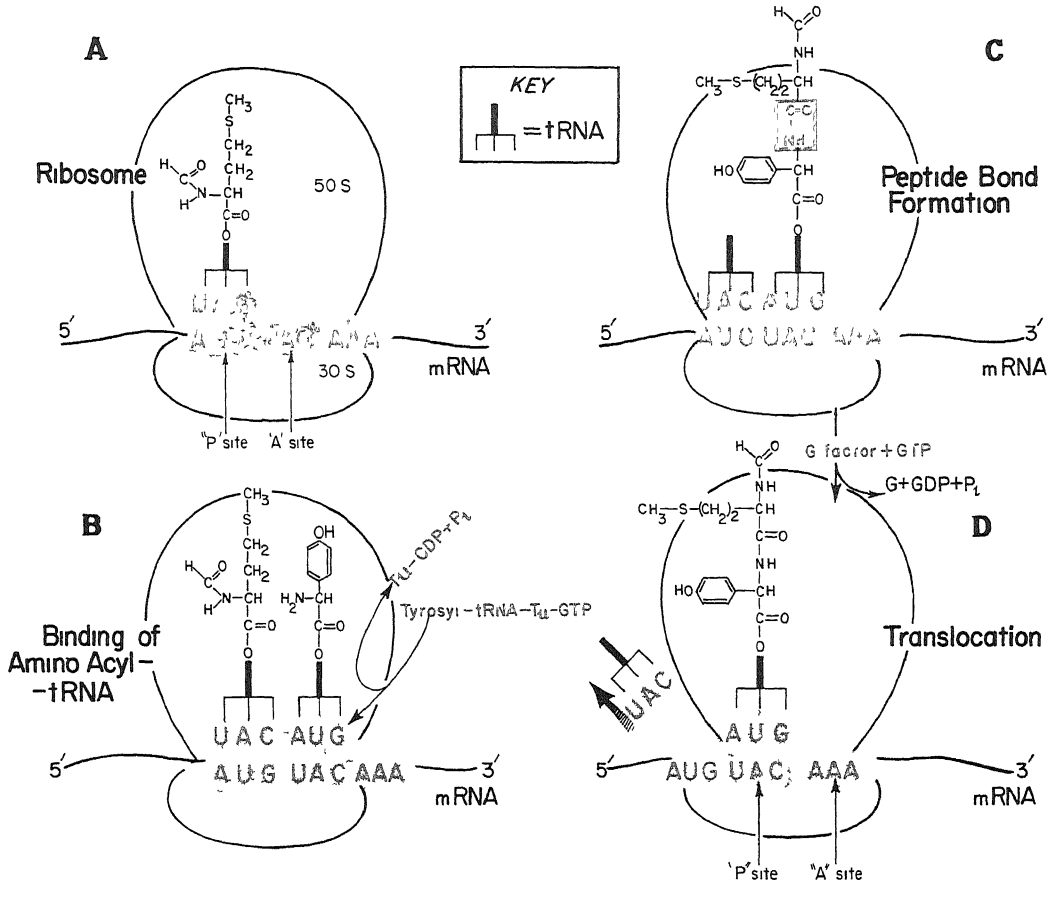


Fig 1 160 Steps in peptide bond formation (see text)

enzyme is that the first site cannot, but the second site can bind an aa-tRNA with a free amino group It is via this specificity of the enzyme that one can reasonably explain the special need for an initiator tRNA with a blocked (formylated)α amino group After peptide bond formation the peptidyl-tRNA is attached to the "A" site of the ribosome-mRNA complex (Fig 1 160*C*)

Step Three. Translocation. In this step the main event is the translocation of the peptidyl-tRNA from the "A" to the "P" site This event requires the *translocation factor* (*translocase*) and GTP During this step, a GTP molecule is cleaved to GDP and P_i At the end of this step the peptidyl tRNA is at the "P" site and the mRNA has moved (relative to the ribosome) by one codon (Fig 1 160*D*) The "A" site, occupied now by the codon to be decoded next, is empty and ready to accept another amino acyl-tRNA as in step one *The translocase* from *E coli* is designated also as the G-factor and the enzyme from rat liver is designated as *transferase II*

PEPTIDE CHAIN TERMINATION

When the synthesis of a peptide chain is completed, the peptide is released from its attachment to the tRNA and to the ribosome-mRNA complex Chain termination is signaled by the nonsense codons UAG, UGA, and UAA The termination process is effected by the protein *termination factors* R_1 and R_2 Factor R_1 effects chain termination signaled by the UAA or UAG codon, whereas factor R_2 elicits the termination response with codons UAA and UGA

INHIBITION OF PROTEIN SYNTHESIS BY ANTIBIOTICS

Several antibiotics inhibit protein synthesis by interfering with steps of this process The

antibiotic chloramphenicol binds to the peptidyl transferase and hinders sterically the binding of the CCA terminus of aa-tRNA to the enzyme The result is inhibition of peptide bond formation and consequently of protein synthesis This antibiotic affects only bacteria ribosomes, not 80S eukaryotic ribosomes

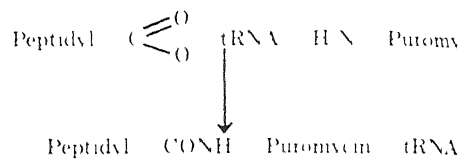
Sparsomycin binds also to peptidyl transferase in a site overlapping that of the chloramphenicol site, and thereby inhibits protein synthesis in an analogous manner This antibiotic affects both bacterial and eukaryotic ribosomes

Cycloheximide (actidione) acts similarly but affects only eukaryotic 80S and not bacterial ribosomes

Streptomycin has two effects on protein synthesis *a*) it inhibits the peptide elongation process, and *b*) it causes *misreading* on certain codons Thus, for example, the antibiotic induces the misreading of codon UUU (phenylalanine) to be read as AUU (isoleucine) The streptomycin effects on protein synthesis are due to the binding of the antibiotic to the 30S ribosomal subunit This binding occurs close to the "A" site and alters the conformation of the site This conformational alteration results not only in inhibition of protein synthesis but also in misreading of codons

Streptomycin Resistance of certain bacterial strains is due to mutations at the streptomycin locus of the genome These mutations cause changes in the 30S ribosomal subunits making them resistant to streptomycin As a result the effect of the antibiotic is greatly reduced It has been recently demonstrated that the component controlled by the streptomycin locus is one of the proteins of the 30S subunit This protein is involved in the fidelity of translation of the genetic message These observations point out the fact that ribosomes participate in protein synthesis not only by providing the necessary surfaces for attachment and binding of the mRNA, tRNAs, and enzymes, but also by influencing the fidelity of translation

Puromycin, another antibiotic, has a structure similar to that of the terminal amino acyl-AMP residue of amino acyl tRNA It therefore binds to the peptidyl transferase next to the peptidyl tRNA The enzyme then catalyzes the formation of a peptide bond between the amino group of puromycin and the activated carboxyl group of the peptidyl-tRNA



The growing peptide is thereby released from the ribosome Thus puromycin elicits premature termination of the growing peptide chain The incomplete polypeptide chain ends in a puromycin residue

POLYRIBOSOMES

After peptide chain initiation the translation of the genetic message proceeds with concomitant movement of the mRNA relative to the ribosome As a result, the initiation end of the mRNA becomes vacant and ready to accept another ribosome and so on This allows a given mRNA to accomodate several ribosomes at the same time Each ribosome constitutes a condensing site through which the mRNA molecule moves with successive exposure of codons and at which a polypeptide chain is growing (Fig 1 161) Such ribosome mRNA aggregates have been isolated and are called *polyribosomes* or *polysomes* A polyribosome can contain a few or many ribosomes The number of ribosomes per polyribosome depends primarily on the length of the messenger RNA strand The longer the mRNA is, the more ribosomes can be accomodated Thus one mRNA strand functions as a template for the synthesis of several molecules of a polypeptide chain at the same time Obviously, at any moment the molecules of the polypeptide being synthesized along the length of the mRNA are shortest near the initiation end, and are close to completion near the termination end (Fig 1 161) Thus the polyribosome arrangement increases the efficiency of utilization of the mRNA molecules A polyribosome functions continuously with ribosomes attaching at the initiation end and ribosomes leaving at the termination end (Fig 1 161) The mRNA serves as a template until its breakdown by specific nucleases This breakdown begins at the initiating end of mRNA Thus, existing ribosomes on the messenger are not interfered with and the translation of the message is completed However, initiation of further polypeptide chain assembly is prohibited

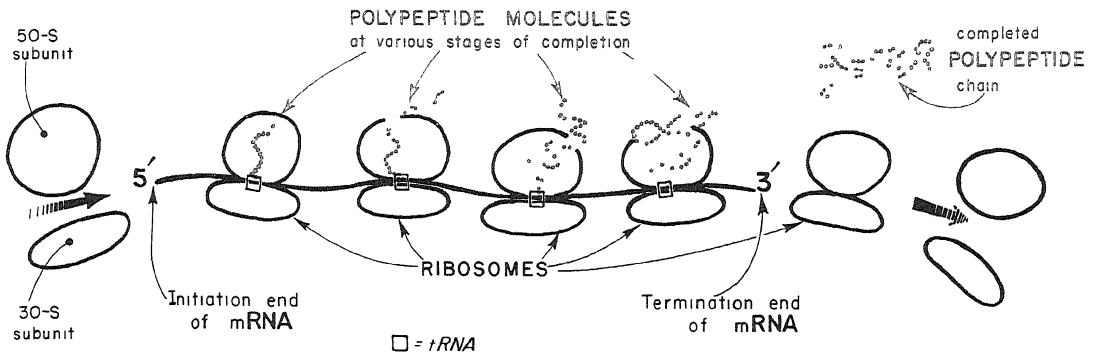


Fig 1 161 Structure and function of a polyribosome (see text)

ASSEMBLY OF POLYPEPTIDE CHAINS INTO PROTEINS

If a gene controls the primary structure of a given polypeptide chain, and if the conformation of the polypeptide chain, viz , its specific folding, is determined by its amino acid sequences (see also chap 5), obviously then the genetic control is extended not only to the primary structure of the polypeptide but also to its specific native functional conformation This folding takes place simultaneously with the synthesis of the polypeptide on the ribosomes Some proteins consist of several chains (subunits), and the assembly of these subunits takes place after the completion of their synthesis On this theme there may be several variations, the most important of which are the following

1) The protein consists of one polypeptide chain In this case the meaning of polypeptide and protein are synonymous

2) The protein derives from the synthesized polypeptide chain by chemical modification of this chain For example, insulin derives from proinsulin by enzymatic cleavage (Fig 1 162*A*)

3) The protein consists of more than one polypeptide chain The assembly of the subunits to form the protein may be accomplished through covalent disulfide bonds as in the case of γ-globulins (Fig 1 162*B*) Most of the time, however, the assembly of the subunits of a protein involves noncovalent forces (Fig 1 162*C*) This is because the subunits "fit" so that their arrangement has minimum free energy

Very often a nonprotein moiety binds to the protein to give a functional unit For example, the enzyme tryptophan synthetase is a complex of pyridoxal phosphate and the protein (Fig 1 162*C*)

The Molecular Basis of Mutations

Any alteration of the base sequence in a gene represents a mutation Such alterations are derived from transitions, transversions, deletions, or insertions of nucleotides within the base sequences of the gene These changes can occur spontaneously or can be induced by mutagenic agents such as x-rays, radioactivity, ultraviolet light, chemical mutagens, etc

MUTATIONS DUE TO TRANSITIONS OR TRANSVERSIONS

A *transition* is a change of one purine base into the other, or one pyrimidine base into the other, a *transversion* is a change of a purine base to a pyrimidine base and vice versa For example

Transition	A↔G,	C↔T (U)
Transversion	A↔C	A↔T (U)
Transversion	G↔C	G↔T (U)

The replacement of one nucleotide by another results in the change of one codon to another Such a change leads to one of the following three possibilities *a*) the codon produced still codes for the same amino acid (*silent mutation*), *b*) the codon produced codes for a different amino acid (*missense mutation*), and *c*) the codon produced is a termination codon (*nonsense mutation*) In the case of missense mutation, the resulting amino acid replacement could lead to a modification of the function of the protein or even complete loss of

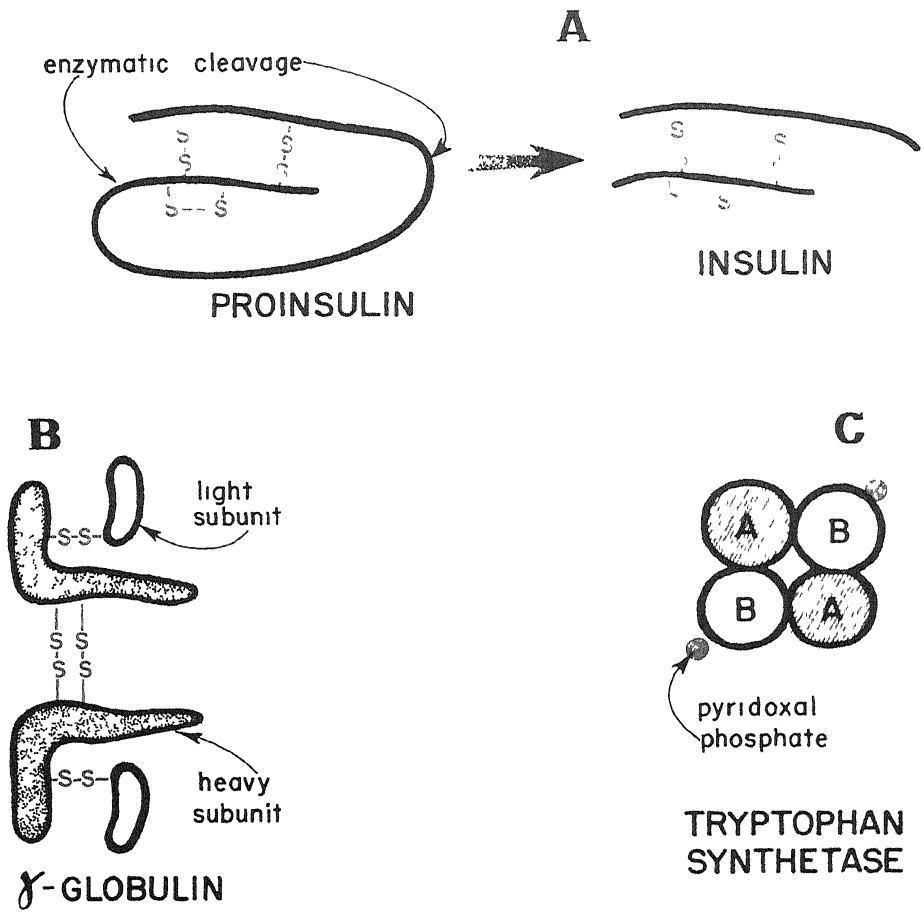


Fig 1 162 Synthesis and assembly of proteins (see text) *A* conversion of proinsulin to insulin *B* a γ globulin molecule consists of four subunits two heavy and two light These subunits are held together by covalent disulfide bridges *C* tryptophan synthetase This enzyme consists of two pairs of subunits *A* and *B* in addition a molecule of pyridoxal phosphate is bound to each B subunit The subunits are held together by noncovalent forces

its activity In the case of a nonsense mutation, the termination codon produced somewhere within the gene will trigger the premature termination of the translation process and result in the formation of an incomplete polypeptide chain

The molecular mechanisms producing transitions can be understood on the basis of improper base pairing In DNA, A pairs always with T, and G pairs with C This pairing occurs with the bases in the *keto form* The bases can undergo *tautomerization* from the keto to the *enol* form When the bases are in the *enol* state their pairing characteristics change Thus, for example adenine in the enol state could pair with cytosine Under physiological conditions, the keto form is predominant and tautomerization to the enol form is rare Watson and Crick (1953) suggested that spontaneous mutations might result from such tautomeric shifts The occurrence, of course, of this type of spontaneous mutations is extremely low Mutations due to improper pairing can be induced by chemical mutagens such as 5-bromouracil (BU) and 2-aminopurine (AP) These *base analogues*, because of their structural similarity to the regular DNA bases, can be incorporated into DNA during replication The analogues, however, undergo frequent tautomerization from the keto to the enol form, and thus BU for example, could pair with either A or G In this manner BU may be correctly incorporated opposite A during DNA replication, but during some later round of replication may pair with G because of tautomerization Ultimately this will result in the replacement of an A-T pair by

a G-C pair as follows

original
A-T pair
A G-C pair
replaces the
original A-T pair
Transition

Alternatively BU may be mistakenly incorporated opposite G but in the next replication may pair correctly with A This will produce an A-T pair at a site formerly occupied by a G-C pair

An A-T pair
replaces the
original G-C pair
transition
Original
G-C pair

Several other mutagens act by altering the bases chemically Monofunctional alkylating agents act by alkylation of bases Ethyl ethanesulfonate for example alkylates guanine residues at position 7 This alkylation increases the rate of tautomerization and causes G to pair occasionally with T Other agents, such as HNO_2 and hydroxylamine, oxidatively deaminate the bases and thus change their pairing properties For example, HNO_2 converts adenine to hypoxanthine which behaves like G, pairing with C The mechanisms responsible for the induction of transversion are not well understood However, transversions are known to occur spontaneously in phages Also many of the abnormal human hemoglobins involve amino acid replacements which can be explained by a single base transversion

READING FRAME SHIFT MUTATIONS INSERTION OR DELETION

These mutations result from *insertion* or *deletion* of one base Consider, for example, the following nucleotide sequence· AAA-AGU-CCA-UCA-G The codons of this segment are read as underlined If, however, a deletion of the A at position 4 occurred, the message read would then be AAA-GUC-CAU-CAG- Note that after codon AAA there is a change in the *reading frame* These kinds of mutations lead obviously to gross structural changes in the coded proteins Deletion of one nucleotide followed by insertion of one nucleotide in the same gene constitutes a *double frame shift* mutation Clearly, in the nucleotide segment between the deletion and insertion, the reading frame is still out of phase However, after the point of insertion, the reading frame is restored to the original Double frame shift mutations have been observed in the enzyme, lysozyme, of the T4 phage

Original sequence -Lys-Ser-Pro-Ser-Leu-Asn-Ala

Double frame shift-mutation -Lys-Val-His-His-Leu-Met-Ala

deletion insertion

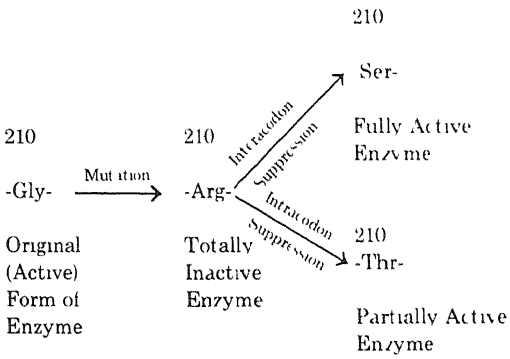
Thus, deletion of one nucleotide from the codon for serine changes the reading frame up to asparagine Insertion of one nucleotide prior to the codon for alanine restores the reading frame

Reading frame shift mutation can be produced in phages by dyes of the acridine series (proflavin, 5-amino-acridine) These dyes can intercalate noncovalently between the planar rings of the bases of DNA The intercalation produces a doubling of the distance between successive bases of DNA As a consequence, during replication an extra nucleotide may be incorporated to match the acridine molecule Thus, an extra nucleotide is inserted into the new strand Subsequent replication of the strand will result in a double strand with an extra base pair During replication a molecule of the dye could similarly interpose itself between the bases of the growing chain in such a way that it matches a nucleotide of the template This can cause the replication enzyme to skip that nucleotide The result is, of course, deletion of one nucleotide Deletions and insertions could also arise by mispairing during *crossing over* and *recombination* of genes

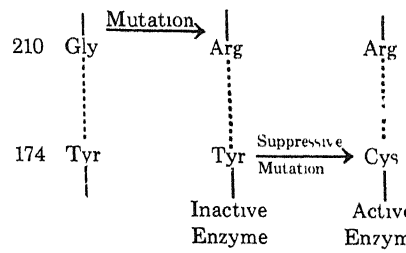
REVERSION AND SUPPRESSION

True reversion of a mutation is due to a correction of the nucleotide change back to the original sequences *Suppression* of a mutation is due to a new mutation, which compensates phenotypically for the original one There are two main types of mutations causing suppressions, intragenic and extragenic *Intragenic* suppressions involve a mutation within the originally mutated gene, whereas *extragenic* suppression involves mutations in a different gene An intragenic suppression could involve an *intracodon compensation* This compensa-

tion could lead to an amino acid replacement which is less detrimental to the protein function For example, mutational replacement of the glycine by arginine at position 210 in tryptophan synthetase of *E coli* leads to an inactive enzyme Further *intracodon mutation* could result in substitution of the arginine by serine (or threonine) The enzyme is now fully active, and thus there is compensation for the effect of the original mutation



Intragenic suppression of a reading frame shift mutation can be compensated for by a second deletion or insertion which restores the reading frame except for the segment between the two mutations Finally, intragenic suppression can involve a second amino acid substitution at some distance from the original change For example, the mutation Gly → Arg at position 210 of tryptophan synthetase results as we have seen in an inactive enzyme A new mutation Tyr → Cys at position 174 produces an active mutant enzyme



Extragenic suppression involves mutation in the anticodon of a tRNA Nonsense mutations, for example, due to formation of the termination codon, UAG, can be suppressed by a mutation in the gene coding for the tyrosine tRNA This mutation changes the anticodon of tRNA tyrosine from ^{3}AUG^{5} to ^{3}AUC^{5} The changed anticodon now matches the nonsense codon

Anticodon tRNA-tyrosine $(^{3}AUG^{5}) \xrightarrow[\text{Mutation}]{\text{Suppressive}} (^{3}AUC^{5})$ $5'—UAG—3'$

Therefore, instead of premature termination of the translation process due to the AUG codon, the synthesis of the polypeptide continues and a tyrosine is inserted in the growing polypeptide chain The change of the tRNA for tyrosine raises the obvious question of how the tyrosine codons are continued to be read Clarification of that point came with the discovery that two different genes code for two tyrosine tRNAs which have the same anticodon One of the two tyrosine tRNAs (*major*) occurs in large amounts compared to the other tRNA (*minor*) The suppression mutation occurs only in the minor tyrosine tRNA

A nonsense suppression of the UAG and UGA codons does not interfere with the normal growth of the cell, presumably because it does not prevent the regular polypeptide termination process When the sequences for the coat protein of the RNA phage R17 were worked out, it was noted that termination was signalled by two successive stop codons UAAUAG This observation prompted the speculation that termination may be signalled by *two stop codons* If this is the case, it may explain why nonsense suppressions due to UAG and UGA codons have no effect on regular chain termination Nonsense suppression of the UAA type interferes with cell growth Whether this interference operates at the level of the normal termination process is not known at the present time Suppressions of missense mutations have also been observed This type of suppression again involves a mutant tRNA

Regulation of Gene Activity

Encoded in the genetic material of every organism is the information necessary for the production of each polypeptide chain (enzyme, structural protein, etc) that will be needed at various stages of development and under various environmental conditions Even the chromosome of a relatively simple organism, the bacterium *E coli*, codes for 2000 to 4000 different polypeptide chains, and mammalian chromosomes contain very much more information By regulating gene activity, a cell avoids a wasteful production of proteins, con-

fines its synthetic activity to the assembly of those proteins that are needed at a given time, and coordinates its metabolic activities in accord with the external environment In fact, single genes or whole sets of genes can be "turned on and off" at the transcriptional level as required during differentiation or environmental changes Cells can also regulate the *rate of gene transcription* and exert certain controls at the *level of translation* Finally, the activity of the proteins themselves in the cell can be regulated by interaction with small molecules or by actual enzymatic modification The discussion that follows is primarily concerned with microbial systems of regulation of transcription, translation, and protein function It is in these relatively simple systems that basic mechanisms have been most elaborately defined While problems of regulation in higher organisms will certainly involve a greater degree of complexity and additional factors, the fundamental concepts elucidated in microorganisms provide a foundation for approaching and ultimately solving such problems

TRANSCRIPTIONAL CONTROLS

Induction and Repression: The Operon Many enzymes, the *constitutive enzymes*, are produced by cells at all times The synthesis of other enzymes is under control, and they are said to be either inducible or repressible *Repressible* enzymes—those commonly involved in biosynthesis—are synthesized *except* when some specific substance (often the product of the metabolic pathway in which the enzyme functions) is present For example, production of the enzymes for tryptophan biosynthesis is "turned off" (repressed) by tryptophan *Inducible* enzymes, which are commonly catabolic enzymes, are *not* synthesized except when the inducer (often the substrate for the enzyme) is present For example, the enzyme β-galactosidase which hydrolyzes lactose and other β-galactosides is synthesized in response to the presence of such compounds in the medium In the absence of inducer, *E coli* cells contain only 1 to 10 units of β-galactosidase per mg of cell protein, whereas in the presence of inducer, the levels of this enzyme reach 10,000 units per mg of protein This increase represents de novo enzyme synthesis rather than activation of a pre-existing enzyme

Jacob and Monod (1961) devised a general model to explain how the production of groups of enzymes that are related in function can be controlled coordinately The genes which code for functionally related enzymes, the *structural genes*, are often linked together on the chromosome Structural genes for the three enzymes that are induced by lactose and other β-galactosides, for example, are linked on the genome of *E coli* They are the *z* gene codes for β-galactosidase, the *y* gene for a permease that facilitates entry of substrate into the cell, and the *a* gene for galactoside acetylase (Fig 1 163) Involved in the control of transcription of these three structural genes are two other genetic elements, the *operator* and the *regulatory gene* As shown in Figure 1 163, the structural genes are directly linked to the DNA segment of the operator The operator and its contiguous structural genes are controlled by the product of the regulatory gene The regulatory gene, the operator, and the linked structural genes constitute a unit of regulation called the *operon* The product of the regulatory gene is a protein called the *repressor* The repressor is produced continuously by the regulatory gene and binds to the operator This turns off transcription of the structural genes of the operon (Fig 1 163*A*) When a β-galactoside (*inducer*) is present, a complex between the inducer and repressor molecules is formed The repressor no longer binds to the operator, and transcription of the structural genes can then occur (Fig 1 163*B*) These controls bring about the synthesis of the enzymatic machinery for lactose metabolism only when actually needed, i e , in the presence of lactose

Among the many kinds of evidence for the Jacob-Monod "Operon Theory" is that obtained from the study of mutants As the model predicts, certain mutations in the regulatory gene or in the operator region lead to constitutive production of the enzymes of the operon For example, a mutation in the regulatory gene could result in mutant repressor molecules that can no longer bind to the operator Such a mutant would have a free operator and hence be able to transcribe the operon even in the absence of inducer (Fig 1 164) Recently, crucial evidence in support of the operon model was provided by the purification of the repressor protein for several operons, including that for the lactose operon by Gilbert and Muller-

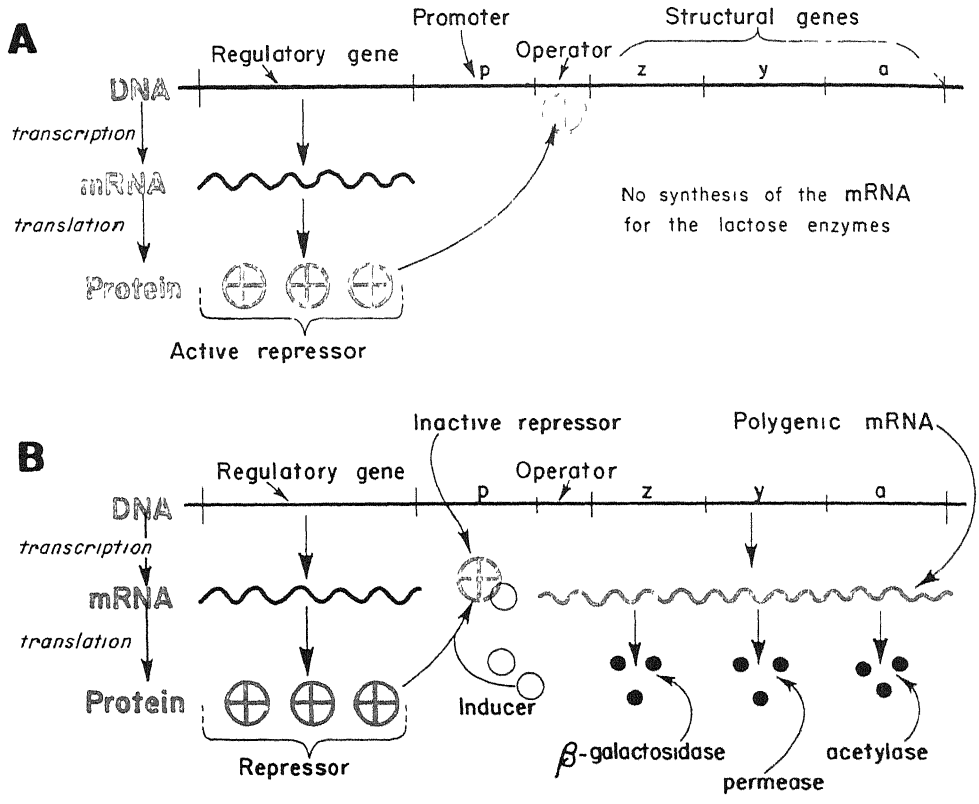


Fig 1 163 The lactose operon of *E coli* in the absence and presence of inducer (see text) *A inducer absent* In this case the repressor binds to the operator and as a result the transcription of the structural genes is blocked *B, inducer present* In this case the inducer binds to the repressor rendering it inactive As a result the structural genes are open for transcription

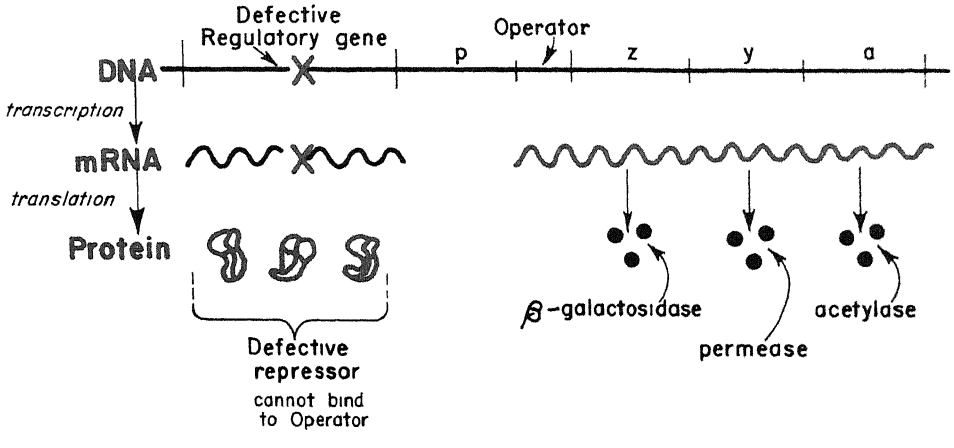


Fig 1 164 Effect of a mutation in the regulatory gene on the function of the lactose operon in *E coli* The mutant repressor cannot bind to the operator and as a consequence the system is open for transcription at all times

Hill (1966) As predicted, the repressor has binding sites for both DNA and inducer Moreover, purified repressor inhibits the transcription of the lactose operon in an in vitro system

The operon model can be used to explain repressible as well as inducible enzyme systems For example, the genes for the five polypeptide chains required for tryptophan

biosynthesis are linked to each other and to an operator region on the *E coli* chromosome (Fig 1 165) There is also a regulatory gene which produces a repressor However, the repressor is *inactive* unless another small molecule, the *corepressor*, is present to change the repressor into its active form Tryptophan or its product acts as the corepressor for the tryptophan operon, thus turning off transcription of the enzymes for its own synthesis, a clearly economical control

In the two kinds of regulatory models described above repressors function—either alone or in interaction with a corepressor—to prevent transcription of an operon, this is called *negative control* There are some groups of enzymes whose operons are apparently subject to *positive control* Such a system is exemplified by the operon controlling the enzymes for arabinose metabolism in *E coli* This operon contains three structural genes coding for three enzymes, an operator region, and the C gene which is a regulatory gene (Fig 1 166) As with the lactose operon, an inducer (in this case, arabinose) is required for transcription of the operon Genetic analysis reveals, however, that unlike the case of the lactose repressor, the C gene product is *required* for induction It is hypothesized that the "C" protein forms a complex with arabinose Interaction of this complex at the operator is required for transcription of the operon

Genetic analysis of mutants has revealed another important regulatory region, the *promoter* In the lactose operon, for example, the region of the promoter (P) lies next to the operator on the other side from the structural genes (Fig 1 163) Promoters are signals along the DNA molecule recognized by the RNA polymerase As a consequence, the enzyme binds to the promoter Such a binding of the enzyme on the promoter is a prerequisite condition for commencement of the transcription process Since binding of the polymerase to the

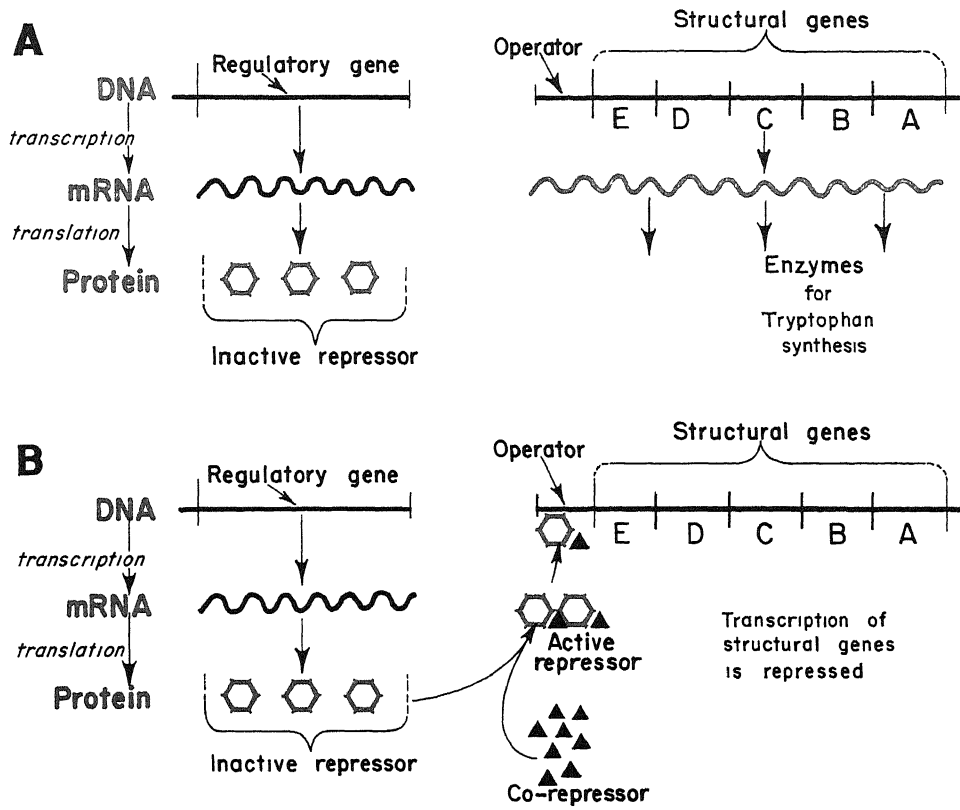


Fig 1 165 The tryptophan operon in *E coli* in the absence and presence of co-repressor (see text) *A corepressor absent* In this case the repressor synthesized by the regulatory gene is inactive and as a result the structural genes are open for transcription *B corepressor present* In this case the corepressor binds to the repressor to produce an active component which then binds to the operator thus transcription is blocked

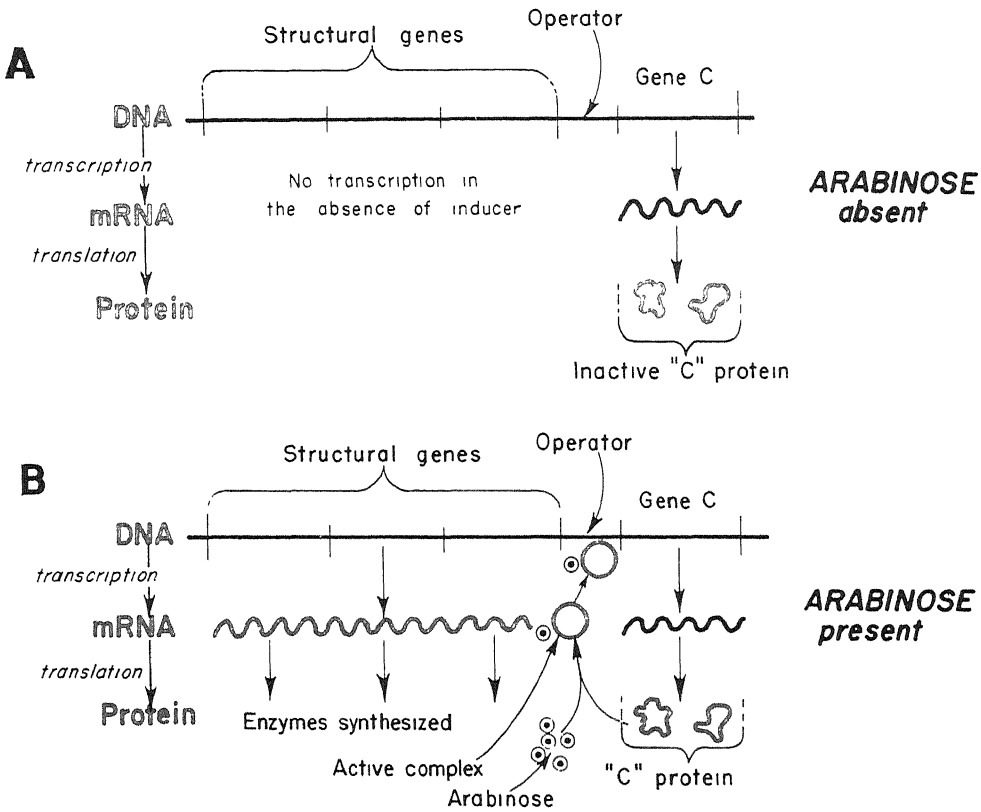


Fig 1 166 Positive control the arabinose operon in *E coli A arabinose absent* In this case the inactive C protein cannot promote transcription of structural genes *B arabinose present* In this case the arabinose forms an active complex with protein C which promotes transcription of structural genes

promoter is the rate-limiting step of the transcription process, mutations in the promoter region affect the *rate of transcription*

A number of regulatory mechanisms appear to act via control of transcription at the promoter For example, in both bacteria and higher organisms the synthesis of many catabolic enzymes is sensitive to the presence of glucose Since glucose is an excellent source of intermediary catabolites and energy for most organisms, it is economical for cells to repress the wasteful synthesis of enzymes for other substrates when glucose is present This phenomenon was given the name *catabolite repression* Catabolite repression in bacteria involves cyclic adenosine monophosphate (cyclic AMP) and a protein which complexes with cyclic AMP The protein is called *CAP or crp protein* The cyclic AMP-CAP complex is a positive control element which is required for proper binding of RNA polymerase to the promoters of certain operons These are the operons that are subject to catabolic repression Such operons require cyclic AMP and CAP for transcription even in the presence of inducer Clearly, lowering of cyclic AMP levels or mutation of the CAP protein would prevent normal transcription Glucose causes catabolic repression by lowering the levels of cyclic AMP within the cell As would therefore be expected, the repressive effect of glucose can be overcome by the addition of large amounts of cyclic AMP

Control by RNA Polymerase. As noted earlier, bacterial RNA polymerase is compounded of several subunits designated as β', β, α, ω, and σ The subunits aggregate specifically to form the holoenzyme, β', β, α_2, ω, σ Strand selection and binding of RNA polymerase to a DNA template is effected through promoters along the DNA molecule The promoters are recognized by the σ-factor of RNA polymerase Once binding of the enzyme to the promoter region is effected, the polymerase will tran-

scribe genes continuously until a *stop signal* on DNA induces the termination of the polymerization process

There are many promoters within a given genome Even the relatively simple T4 phage appears to possess several promoters The intriguing possibility exists that there are differences between promoters and that different promoters require different σ-factors for recognition The "catabolic sensitivity" of some promoters (see above) suggests that such differences exist, as does the discovery that σ-factors are involved in the coordinated expression of the genome of the T4 phage, a phage which contains several genes The proteins coded by these genes are needed at various times during the life cycle of the phage inside the host cell For example, the enzyme lysozyme is needed at the very end of the life cycle of phage, to produce the lysis of the host cell, and thus liberate the viral particles It would be detrimental to viral replication if the gene for lysozyme were transcribed immediately after injection of the phage into the host cell since it would produce a premature lysis of that cell

After the DNA of T4 phage is injected into an *E coli* cell, the various viral genes are expressed in a coordinate fashion as follows the "early" genes are transcribed first, then the "delayed early" genes, and finally the late genes The early genes are transcribed by the *E coli* RNA polymerase Presumably, the promoter(s) of the early genes are recognized by the σ-factor of the host cell However, the promoter(s) of the delayed early and late genes cannot be recognized by the host σ-factors Indeed, one of the early genes codes for a specific σ-factor which recognizes the promoter(s) of the delayed early genes but not of the late genes In turn, one of the delayed early genes codes for a second viral σ-factor specific for the promoters of the late genes

In the case of transcription of the late genes of T7 phage a specific T7 polymerase is needed The enzyme appears to be a linear end-to-end protein (no subunits) of 110,000 MW The T7 polymerase does not bind *E coli* holoenzyme or core enzyme and is insensitive to rifamycin The enzyme is highly active with the T7 DNA, but it does not transcribe other DNA templates

Another kind of control at the transcriptional level is exerted by modification of the RNA polymerase core enzyme Recent studies on bacterial sporulation suggest that the transition from the vegetative form to the spore involves changes in RNA polymerase Thus, a modified polymerase is found in *Bacillus subtilis* at the onset of sporulation The change of enzyme is due to a modification of the β-subunit The modified enzyme does not contain σ-factor, presumably because it cannot bind it Thus, the cessation of the transcription of the genes expressed in the vegetative state coincides with the "inactivation" of the vegetative enzyme (negative control) Activation of the genes expressed at the spore stage is effected by a new σ-factor Whether the new σ-factor binds to the modified vegetative enzyme or to an altogether new polymerase enzyme is not yet known

Therefore, *specific promoters*, *sigma factors*, and *changes in the "core" enzyme* can all contribute to the coordinated regulation of genes at the transcriptional level

It has also been suggested that many operons may be under the control of further positive control elements which react with the polymerase to change its ability to initiate transcription at certain promoters Indeed, there is now evidence for a *psi factor* which exerts specific positive control over the transcription of ribosomal RNAs Moreover, there is a small molecule, guanosine tetraphosphate (ppGpp), whose accumulation inhibits the psi factor, and thus decreases the production of ribosomal and transfer RNAs

Finally, in addition to the protein factors, sigma, and CAP mentioned above, there is a protein factor required for termination of transcription in certain in vitro systems This factor, the *rho factor*, may well have a regulatory function or at least be involved in the normal process of transcription in vivo

TRANSLATIONAL CONTROLS

In addition to controls of gene activity at the transcriptional level, gene expression may also be regulated by controls acting at the translational level These controls ultimately affect the amount of a given protein produced by its messenger RNA, thus inhibition of translation of a given messenger or changes in the rate of its translation lead to changes in the amount of the protein coded by this messenger For example, the RNA phage R17 codes for three proteins the attachment protein, the coat protein

and the RNA replicase protein necessary for viral replication (Fig 1 167) The production of the coat protein is about 10 times greater than the production of the other two proteins This difference is due to the greater affinity of ribosomes for the initiating signal of the coat protein than for the attachment protein The conformation of the initiating site of the RNA replicase gene makes it unavailable for ribosomes binding, the site becomes accessible to ribosomes only when the ribosomes moving across the coat protein gene induce a change in its conformation (Fig 1 167) Thus, translation of the replicase gene depends on translation of the coat protein gene Therefore, the affinity of ribosomes toward initiating sites of mRNAs can be an important determinant of the rate of translation of a given gene

Another very important control device is exerted at the translational level by specific repressor proteins Such repressors bind to mRNA and inhibit translation For example, in the case of phage R17 translation, the synthesized coat protein subunits are specific repressors of the RNA replicase gene The synthesis of this enzyme stops The replicase already synthesized serves for the viral replication, and thus a wasteful synthesis is avoided Control by repressors acting at the translational level may be of great importance for eukaryotic systems (see "Tomkins theory" below)

There are many possibilities for control other than those already described Control by breakdown of mRNA, for example, clearly occurs at least in some systems The actual nuclease which hydrolyzes mRNA in vivo is unknown, however, recent experiments with *E coli* have identified the gene specifying this enzyme Elucidation of the nature of the regulation of mRNA degradation should prove to be of great interest

Finally there are innumerable controls related to the activity of proteins themselves (*post-translational controls*) Some of the most

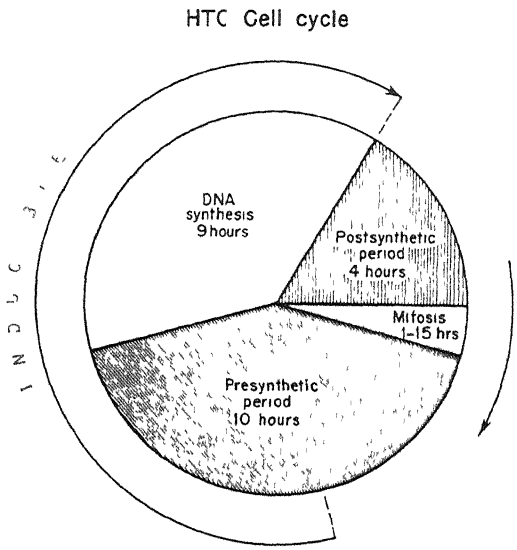


Fig 1 168 The cell cycle of rat hepatoma cells in tissue culture (*HTC*) showing periods during which tyrosine amino transferase can be induced by adrenal steroids Redrawn from Tomkins et al 1969

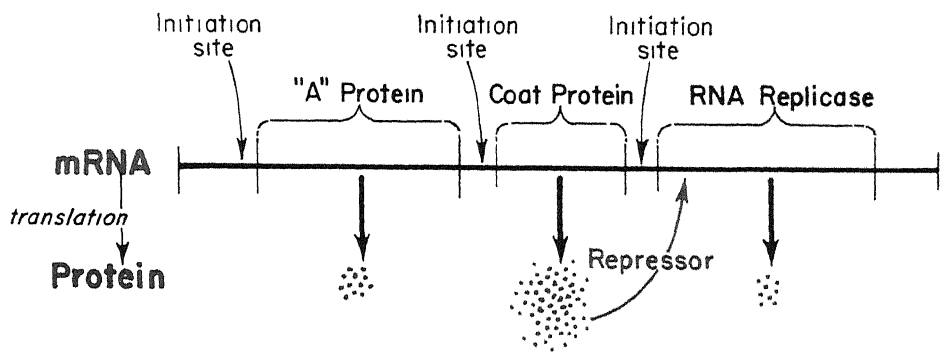


Fig 1 167 Single stranded RNA of phage R17 a polygenic messenger RNA The rate of translation of the coat protein gene is greater than that of the A-protein and the replicase because the conformation at the initiation site of that gene facilitates the attachment of ribosomes much more readily than the conformation of the other two sites In fact the conformation of the initiating site of replicase renders this gene unavailable for translation and it is only when the ribosomes move across the coat protein approaching the replicase region that an appropriate conformational change is induced in the latter region rendering it available for translation Furthermore the coat protein is a specific repressor of the replicase gene inasmuch as it binds to that gene and thereby blocks its continued translation

important of these include feedback inhibition of enzyme function by allosteric mechanism, and actual modification of proteins by phosphorylation, adenylation, or limited proteolysis

SOME COMMENTS ON CONTROL IN HIGHER ORGANISMS

Fundamental differences between regulation in bacterial (*prokaryotic*) and higher (*eukaryotic*) cells are likely to result from the important differences in cellular organization The elaborate organization of the chromosomes in eukaryotic cells poses certain problems We do not know, for example, what regulatory role may be played by the *histones*, the highly basic proteins to which much of the DNA is bound We also do not know the mechanism for selecting some of the many mRNAs that are transcribed in the nucleus to be transported into the cytoplasm for translation or even the mechanism of this transport

A variety of technical advances offers the promise of facilitating experiments on regulation in mammalian cells hitherto regarded as "impossible " For example, it is now possible to cause *fusion* between cells of different types and different activities, by means of certain viruses (usually Sendai virus) Experiments with fused cells have already provided indications that positive control elements are involved in both DNA and RNA synthesis in mammalian cells The general increase in the availability of tissue culture cell lines which are capable of specific synthetic activities in vitro has opened up broad new areas of study As a case in point, Tomkins and his colleagues (1969) have studied the effect of adrenal steroids on the synthesis of the enzyme tyrosine aminotransferase in cultured rat hepatoma cells (HTC cells) This enzyme is induced by steroids, but only during certain periods of the cell cycle (Fig 1 168) Tomkins has proposed a theory of induction of enzymes in mammalian cells which involves two genes—a structural gene for the induced enzyme and a regulatory gene The product of the regulatory gene, the repressor, binds to the mRNA of the structural gene and reversibly inhibits its translation Furthermore, it is assumed that the mRNA can only be degraded when its translation is inhibited by the repressor Thus, the repressor has a dual effect *a*) it inhibits messenger translation, and *b*) it promotes messenger degradation According to the theory, in the presence of the inducer the repressor is inactivated This results in mRNA translation, prevention of mRNA degradation, and increase synthesis of the enzyme (induction) Transcription of the structural gene itself also seems to be directly increased by hormone action These events take place only during particular (inducible) periods of the cell cycle During the noninducible phases, the transcription of both structural and regulator genes is repressed The pre-existing messenger is more stable during this phase of the cell cycle and its translation continues Obviously, inducers have no effect in this phase of the cell cycle

The complex regulatory processes of the coordinate and selective expression of the various genes lead not only to harmonious cell function, but also to changes as required during cell differentiation Selective, sequential and coordinate gene expression is achieved, therefore, by events taking place at the transcriptional, translational, and post-translational levels These events are triggered by molecular signals emanating from the internal or external cell milieu In a sense it is the dynamic balance existing between the genes and their environment that ensures their selective, timely, and coordinate expression

BIBLIOGRAPHY

ABBOTT, B C, A V HILL AND J V HOWARTH The positive and negative heat production associated with a nerve impulse *Proc Roy Soc (London) [Biol] Ser B* 148 149-187 1958

ADRIAN, R H W K CHANDLER, AND A L HODGKIN Slow changes in potassium permeability in skeletal muscle *J Physiol (London)* 208 645-668 1970

AUBERT, X Le couplage énergétique de la contraction musculaire In *These D'agregation de L'enseignement Superieur*, edited by U Louvain Bruxelles, Belgium Editions Arscia, 1965

AVERY, O T, C M M MACLEOD, AND M MCCARTY Chemical nature of the transformation factor *J Exp Med* 9 137-158, 1944

BALIS M E Metabolism of oxypurines in man *Advances Clin Chem* 10 157-211 1967

BÁRÁNY, M ATPase activity of myosin correlated with speed of muscle shortening *J Gen Physiol* 208 197-218, 1967

BASMAJIAN J V *Muscles Alive Their Functions Revealed by Electromyography*, Ed 2 Baltimore Williams & Wilkins 1967

BECK W S Deoxyribonucleotide synthesis and the role of vitamin B_{12} in erythropoiesis *Vitamins Hormones (NY)* 26 413-442 1969

BECROFT D M O, L I PHILLIPS, AND A SIMMONDS Hereditary orotic aciduria Long term therapy with uridine and a trial of uracil *J Pediat* 75 855-891 1969

BENNETT H S Modern concepts of structure of striated muscle *Amer J Phys Med* 34 46–67 1955
BENNETT H S The concepts of membrane flow and membranous circulation in mechanisms for active transport and ion pumping, *J Biophys Biochem Cytol* 2 (Suppl.) 99–103 1956
BERNSTEIN J *Elektrobiologie* Braunschweig Germany Friedr Vieweg Sohn 1912
BLAND J H (ed.) Seminars on the Tesch Nahin syndrome *Fed Proc* 27 1027–1112 1968
BROOKHART G L S COOMBS AND J C ECCLES The recording of potentials from motoneurons with an intracellular electrode *J Physiol* 117 431–460 1952
BUCHTHAL F C GUILD AND P ROSENFALCK Multielectrode study of the territory of a motor unit *Acta Physiol Scand* 39 83–104 1957
BUCHTHAL F E KAISER AND P ROSENFALCK The rheology of the cross striated muscle fibre *Biol Medd Dansk Vid Selsk* 21 1–318 1951
CAIRNS J The chromosome of *E coli* *Cold Spring Harbor Symp Quant Biol* 28 43–46 1963
CASKEY C T D MARSHTON AND J B WYNGAARDEN The enzymology of feedback inhibition of glutamine phosphoribosyl pyrophosphate amidotransferase by purine ribonucleotides *J Biol Chem* 239 2570–2579 1964
CHARGAFF E Structure and function of nucleic acid as cell constituents *Fed Proc* 10 654–659 1951
COOMBS J S J C ECCLES AND P FATT The inhibitory suppression of reflex discharge from motoneurons *J Physiol (London)* 130 396–413 1955
CRICK F H C The genetic code *Sci Amer* 215 Sept. 55–62 1966
CSAPO A Molecular structure and function of smooth muscle In *The Structure and Function of Muscle* edited by G H Bourne New York Academic Press 1960 vol 1 chap 8 p 229–264
CURCI M R, AND M D DONACHI An attempt to find pyrimidine inhibitors of a mammalian aspartate transcarbamylase *Biochim Biophys Acta* 85 338–341 1964
CURTIS H J AND K S COLE Transverse electric impedance of the squid giant axon *J Gen Physiol* 21 757–765 1938
DAVIES R E A molecular theory of muscle contraction Calcium dependent contractions with hydrogen bond formation plus ATP dependent extensions of part of the myosin actin cross bridges *Nature (London)* 199 1068–1074 1963
DAVSON H *A Textbook of General Physiology* Ed 3 Boston Little Brown 1964
EBASHI S AND M ENDO Calcium ion and muscle contraction *Progr Biophys* 18 123–183 1968
ECCLES J C *The Neurophysiological Basis of Mind* London Oxford Univ Press 1953
ERLANGER J AND H S GASSER *Electrical Signs of Nervous Activity* Philadelphia Univ Pennsylvania Press 1937
FATT P AND B KATZ Spontaneous subthreshold activity of motor nerve endings *J Physiol (London)* 117 109–128 1952
FAWCETT D *The Cell, an Atlas of Fine Structure* Philadelphia Saunders 1966
FENN W O A quantitative comparison between the energy liberated and the work performed by the isolated sartorius muscle of the frog *J Physiol (London)* 58 175–203 1923
FENN W O The relation between the work performed and the energy liberated in muscular contraction *J Physiol (London)* 58 373–395, 1924
FRANKLIN, R E, AND R GOSLING *Nature*, 171 740 1953
FRANZINI ARMSTRONG C AND K R PORTER Sarcolemmal invaginations constituting the T system in fish muscle fibers *J Cell Biol* 22 675–696, 1964
FRISCH, L (ed.) The genetic code *Cold Spring Harb Symp Quant Biol* 31 1–762 1966
FRISCH, L (ed.) The mechanism of protein synthesis *Cold Spring Harb Symp Quant Biol* 34 1–855, 1969
GASSER, H S AND HILL, A V The dynamics of muscular contraction *Proc Roy Soc (London) Ser B* 96 398–437, 1924
GERHART J C, AND A B PARDEE The enzymology of control by feedback inhibition *J Biol Chem* 237 891–896, 1962
GILBERT, W, AND B MULLER HILL Isolation of the lac repressor *Proc Nat Acad Sci USA* 56 1891 1966
GOULIAN, M, A KORNBERG, AND R L SINSHEIMER Enzymatic synthesis of infectious phage ϕX 174 DNA *Proc Nat Acad Sci USA* 53 84–91 196[illegible]
GUIMAN A B The biological significance of uric acid *Harvey Lect* 60 [illegible] 1965
HAGER S E AND M E JONES A glutamine dependent enzyme for the synthesis of carbamyl phosphate for pyrimidine biosynthesis in rat liver *J Biol Chem* 4 [illegible] 1966
HAKANSSON C H Action potential recorded intra and extracellularly from the isolated frog sartorius fiber in Ringer solution and in air *Acta Physiol Scand* 49 [illegible] 1960
HEINZ E AND P M WALSH Exchange diffusion transport and intracellular level of amino acid in Ehrlich carcinoma cells *J Biol Chem* 236 1488–1491 1961
HERSHKO A C HERSHKO AND E MAGER Increased formation of phosphoribosyl pyrophosphate in red blood cell of one gouty patient *Proc Nat Acad Sci* 49 796–803 1963
HILL A V *Muscular Movement* Ed 2 New York McGraw Hill 1931
HILL A V The heat of shortening and the dynamic constants of muscle *Proc Roy Soc (London) Ser B* 126 136–195 1938
HILL A V The heat production of muscle and nerve *Ann Rev Physiol* 21 1–18 1959
HILL A V *Live and Dead Experiments on Muscle Mechanics* London Cambridge Univ Press 1960
HILL D K The effect of temperature on the resting tension of frog's muscle in hypertonic solution *J Physiol (London)* 108 241–246 1949
HODGKIN A L The ubthreshold potentials in a crustacean nerve fibre *Proc Roy Soc (London) [Biol] Ser B* 126 87–121 1938
HODGKIN A L Ionic movements and electrical activity in giant nerve fibres *Proc Roy Soc (London) [Biol] Ser B* 148 1–37 1957
HODGKIN A L The ionic basis of nervous conduction *Science* 145 1148–1154 1964
HODGKIN A L AND A F HUXLEY Resting and action potentials in single nerve fibre *J Physiol (London)* 104 176–195 1945
HODGKIN A L AND A F HUXLEY Current carried by sodium and potassium ions through the membrane of the giant axon of *Loligo J Physiol (London)* 116 449–472 1952a
HODGKIN A L AND A F HUXLEY The components of membrane conductance in the giant axon of *Loligo J Physiol (London)* 116 473–496 1952b
HODGKIN A L AND A F HUXLEY The dual effect of membrane potential on sodium conductance in the giant axon of *Loligo J Physiol (London)* 116 497–506 1952c
HODGKIN A L AND A F HUXLEY A quantitative description of membrane current and its application to conduction and excitation in nerve *J Physiol (London)* 117 500–544 1952d
HOLLEY R W J APGAR G A EVERETT J T MADISON M MARQUISSEE S H MERRILL J R PENSWICK AND A ZAMIR The base sequence of yeast alanine transfer RNA *Science* 147 1462 1965
HURSH, J B Conduction velocity and diameter of nerve fibers *Amer J Physiol* 127 131–139 1939
HUXLEY A F Electrical processes in nerve conduction In *Ion Transport across Membranes*, edited by H T Clarke and D Nachmansohn New York Academic Press 1954 p 23–34
HUXLEY A F Excitation and conduction in nerve quantitative analysis *Science* 145 1154–1159 1964
HUXLEY H E Electron microscope studies on the structure of natural and synthetic protein filaments from striated muscle *J Molec Biol* 7 281–308, 1963
HUXLEY H E Structural evidence concerning the mechanism of contraction in striated muscle In *Muscle Proceedings of a Symposium Held at the Faculty of Medicine University of Alberta*, edited by W M Paul E E Daniel, C M Kay and G Monckton Oxford Eng Pergamon, 1964 p 3–28
HUXLEY, H E The mechanism of muscular contraction *Science* 164 1356–1365 1969
HUXLEY, H E, AND J HANSON The molecular basis of contraction in cross striated muscles In *The Structure and Function of Muscle*, edited by G H Bourne New York Academic Press 1960, vol 1, chap 7, p 183–227
JACOB, F, AND J MONOD Genetic regulatory mechanisms in the synthesis of proteins *J Mol Biol* 3 318 1961
KELLEY, W N M L GREFNE F M ROSENBLOOM, J F

HENDERSON AND J E SEEGMILLER Hypoxanthine guanine phosphoribosyl transferase deficiency in gout *Ann Intern Med* 70 155–206 1969

KENDREW J C Side chain interactions in myoglobin *Brookhaven Sympos Biol* 15 216–228 1962

KHORANA H G Synthesis in the study of nucleic acids *Biochem J* 109 709–725 1968

KORNBERG A Active center of DNA polymerase *Science* 163 1410 1969

LASKOWSKI M DNases and their use in the studies of primary structure of nucleic acids *Advances Enzym* 29 165–220 1967

LEHNINGER A L *Biochemistry*, Worth New York 1970

LLOYD D P C Facilitation and inhibition of spinal motoneurones *J Neurophysiol* 9 421–438 1946

LOEWI O Uber humoral Ubertragbarkeit der Herznervenwirkung I Mitteilung *Pflueger Arch* 189 239–242 1921

LOW B W AND J T EDSALL Aspects of protein structure In *Currents in Biochemical Research* edited by D E Green New York Interscience 1956 p 378–433

MAGASANIK B The metabolic regulation of purine interconversions and of histidine biosynthesis In *The Chemical Basis of Development* edited by W D McElroy and B Glass Baltimore Johns Hopkins Press 1958 p 485–494

MALAWISTA S E Colchicine a common mechanism for its anti in flammatory and anti mitotic effects *Arthritis Rheum* 11 191–197 1968

McILWAIN H *Biochemistry and the Central Nervous System* Ed 3 Boston Little Brown 1966

MENDEL L B *Nutrition The Chemistry of Life* New Haven Conn Yale Univ Press 1923

MESELSON, M AND F W STAHL The replication of DNA in *E coli* *Proc Nat Acad Sci USA* 44 671 1958

MIESCHER F, 1869

MURRAY A W D C ELLIOTT AND M R ATKINSON Nucleo tide biosynthesis from preformed purines in mammalian cells regula tory mechanisms and biological significance *Progr Nucl Acid Res* 10 87–119, 1970

NAKANISHI S K ITO AND M TATIBANA Two types of car bamyl phosphate synthetase in rat liver chromatographic resolution and immunological distinction *Biochem Biophys Res Commun* 33 774–787 1968

NIRENBERG, M W Genetic memory *J A M A* 206 1973–1977 1968

NIRENBERG, M W AND J H MATTHAEI The dependence of cell free protein synthesis in *E coli* upon naturally occurring or synthetic polyribonucleotides *Proc Nat Acad Sci USA* 47 1588, 1961

NOMURA, M Assembly of bacterial ribosomes *Science* 179 864–873 1973

NOVIKOFF A B AND E HOLTZMAN *Cells and Organelles* New York Holt Rinehart and Winston 1970

PAULING L AND R B COREY Configurations of polypeptide chains with formed orientations around single bonds two new plated sheets *Proc Nat Acad Sci USA* 37 729–740 1951

PEACHEY L D The sarcoplasmic reticulum and transverse tubules of the frog's sartorius *J Cell Biol* 25 209–231 1965

PETERSON I AND E KUGELBERG Duration and form of action potential in the normal human muscle *J Neurol Neurosurg Psychiat* 12 124–128 1949

PHELPS P Polymorphonuclear leukocyte motility *in vitro* Possible release of a chemotactic substance after phagocytosis of urate crystals by polymorphonuclear leukocytes *Arthritis Rheum* 12 197–204 1969

PINSKY, L AND R S KROOTH Studies on the control of pyrimi dine biosynthesis in human diploid cell strains Effect of 6-azauridine on cellular phenotype *Proc Nat Acad Sci USA* 57 925–932, 1967

PINSKY, L AND R S KROOTH Studies on control of pyrimidine biosynthesis in human diploid cell strains Effects of 5 azaorotic acid barbituric acid, and pyrimidine precursors on cellular phenotype *Proc Nat Acad Sci USA* 57 1267–1274 1967

PODOLSKY R J AND L E TEICHHOLZ The relation between calcium and contraction kinetics in skinned muscle fibres *J Physiol (London)* 211 19–35 1970

PORTER K R AND G E PALADE Studies on the endoplasmic re ticulum III Its form and distribution in striated muscle cells *J Biophys Biochem Cytol* 3 269–300 1957

PRAGER M D J E YOUNG AND I C ATKINS A study of the possible role of feedback inhibition of aspartate transcarbamylase in regulation of pyrimidine synthesis in human leukocytes *J Lab Clin Invest* 70 768–776, 1967

PROSSER C L AND F A BROWN JR *Comparative Animal Physiology*, Ed 2 Philadelphia Saunders 1961

RENSHAW B Central effects of centripetal impulses in axons of spinal ventral roots *J Neurophysiol* 9 191–204 1946

ROBERTSON J D Unit membranes A review with recent new studies of experimental alterations and a new subunit structure in synaptic membranes In *Cellular Membranes in Development* ed ited by M Locke New York Academic Press 1964 p 1–81

ROSENBLOOM F M J F HENDERSON AND W N KELLEY A defect of feedback inhibition in adult gout with overproduction of uric acid *Arthritis Rheum* 11 507 1968

SCHNEIDER M F Linear electrical properties of the transverse tu bules and surface membrane of skeletal muscle fibers *J Gen Physiol* 56 640–671 1970

SMITH L H JR M SULLIVAN AND C M HUGULEY JR Py rimidine metabolism in man The enzymatic defect of orotic aci duria *J Clin Invest* 40 656–664 1961

SORENSON L B Hyperuricemia and gout *Advances Intern Med* 15 177–201 1969

TASAKI I *Nerve Excitation a Macromolecular Approach* Spring field Ill Thomas 1968

TAYLOR J H [ed] *Molecular Genetics* New York Academic Press 1963 and 1967 parts I and II

TOMKINS G M D T GELEHURTER D M GRANNER H H SAMUELS, AND G B THOMPSON Control of specific gene ex pression in higher organisms *Science* 166 1474 1969

TOWER, D B Molecular transport across neural and non neural membranes In *Properties of Membranes* edited by D B Tower S A Luse and H Grundfest New York Springer 1962

VERATTI E Ricerche sulla fine struttura della fibra muscolare striata *Arch Ital Biol* 37 449–454 1902

VOET D AND A RICH The crystal structures of purines pyrimi dines and their intermolecular complexes *Progr Nucl Acid Res* 10 183–265 1970

WATSON J D *Molecular Biology of the Gene* W A Benjamin New York 1970

WATSON J D AND F H C CRICK Genetic implications of the structure of deoxyribonucleic acid *Nature* 177 964 1953

WEBER A On the role of calcium in the activity of adenosine 5' tri phosphate hydrolysis by actomyosin *J Biol Chem* 234 2764–2769 1959

WILKINS, M H F AND S ARNOTT *J Molec Biol* 11 391 1965

WYCKOFF H W, K D HARDMAN N M ALLEWELL T INGAMI D TSERNOGLOU L N JOHNSON AND F M RI CHARDS The structure of ribonuclease s at 6A resolution *J Biol Chem* 242 3749–3753 1968

WYNGAARDEN, J B AND D M ASHTON The regulation of ac tivity of phosphoribosylpyrophosphate amidotransferase by purine ribonucleotides a potential feedback control of purine biosynthesis *J Biol Chem* 234 1492–1496, 1959

WYNGAARDEN J B AND R A GREENLAND The inhibition of succinoadenylate kinosynthetase of *Escherichia coli* by adenosine and guanosine 5' monophosphates *J Biol Chem* 238 1054–1057 1963

WYNGAARDEN J B Gout In *The Metabolic Basis of Inherited Disease*, Ed 2 edited by J B Stanbury J B Wyngaarden and D S Fredrickson New York McGraw Hill 1966 p 667–728

YANOFSKY C Gene structure and protein structure *Sci Amer* 216 (Suppl 5) 80–94 1967

YU, T F AND A B GUTMAN Uric acid nephrolothiasis in gout predisposing factors *Ann Intern Med* 67 1133–1148 1967

ZAMECNIK P C Protein synthesis *Harvey Lectures* 54 256 1960

GENERAL READING

ADRIAN E D *The Mechanism of Nervous Action* Philadelphia Univ Pennsylvania Press 1935 (Reprinted 1959)

ADRIAN E D AND D W BRONK The discharge of impulses in motor nerve fibers II The frequency of discharge in reflex and voluntary contractions *J Physiol (London)* 67 119 151 1929

ADRIAN R H The effect of internal and external potassium concentration on the membrane potential of frog muscle *J Physiol (London)* 133 631-658 1956

ADRIAN R H I L COSTANTIN AND L D PEACHEY Radial spread of contraction in frog muscle fibres *J Physiol (London)* 204 231-257 1969

ALLISON J B Biological evaluation of proteins *Physiol Rev* 35 664 700 1955

ALTSCHUL A M (chairman) In *Advances in Chemistry* series 57 Washington D C Amer Chem Soc 1966

AMERICAN CHEMICAL SOCIETY *Advances in Protein Chemistry* series 57 Washington D C Amer Chem Soc , 1966 vol 21

ANFINSEN C *The Molecular Basis of Evolution* New York Wiley 1960

BASMAJIAN J V Control and training of individual motor units *Science* 141 440-441 1963

BASMAJIAN J V M BAEZA AND C FABRIGAR Conscious control and training of individual spinal motor neurons in normal human subject *J New Drugs* 5 78, 1956

BONDY P K (ed) *Diseases of Metabolism* Ed 6 Philadelphia Saunders 1969

BOYER P (ed) *The Enzymes* Ed 3 New York Academic Press 1970 vols I and II

BRACHET J AND A E MIRSKY In *Cells and Their Component Parts* New York Academic Press 1961, vol II

BRODSKY W A A E SHAMOO AND I L SCHWARTZ Dissipative transport processes In *Handbook of Neurochemistry*, edited by A Lajtha New York Plenum, 1971 vol 5

BROOKHAVEN NATIONAL LABORATORY *Protein Structure and Function*, report of a symposium Brookhaven Symposia in Biology 13, 1960

BROOKHAVEN NATIONAL LABORATORY *Subunit Structure of Proteins Biochemical and Genetic Aspects*, report of a symposium Brookhaven Symposia in Biology, 17, 1964

BROOKHAVEN NATIONAL LABORATORY *Structure Function and Evolution in Proteins*, report of a symposium Brookhaven Symposia in Biology, 21 1969 vol I and II

BULLER A J J C ECCLES, AND R M ECCLES Differentiation of fast and slow muscles in the cat hind limb *J Physiol (London)* 150 399-416, 1960

CALDWELL P C A L HODGKIN R D KEYNES AND T L SHAW The effects of injecting "energy rich" phosphate compounds on the active transport of ions in the giant axons of *Loligo J Physiol (London)* 152 561-590, 1960

COLE K S , AND J W Moore Ionic current measurements in the squid giant axon membrane *J Gen Physiol* 44 123-167, 1960

COLLANDER, R Cell membranes Their resistance to penetration and their capacity for transport In *Plant Physiology*, edited by F C Steward New York Academic Press, 1959, vol II

CONNELLY C M Recovery processes and metabolism of nerve In *Biophysical Science*, edited by J L Oncley New York Wiley, 1959 chap 51, p 475-484

CSÁKY T Z Transport through biological membranes *Ann Rev Physiol* 27 415-450, 1965

DOWBEN, R M *General Physiology* New York Harper & Row, 1969

DUPRAW, E J *Cell and Molecular Biology* New York Academic Press, 1968

DU VIGNEAUD, V Interrelationships between choline and other methylated compounds *Biol Symp* 5 234-247, 1941

DU VIGNEAUD, V The significance of labile methyl groups in the diet and their relation to transmethylation *Harvey Lect* XXXIX 39-62, 1943

EDMAN, K A The rising phase of the active state in single skeletal muscle fibers of the frog *Acta Physiol Scand* 79 167-173, 1970

EINSTEIN A *Investigations on the Theory of the Brownian Movement* New York Dover 1926

ELIOT K A C I H PAGE AND J H QUASTEL (eds) *Neurochemistry* Ed 2 Springfield Ill Thomas 1955

FLORKIN M AND E STOTZ (eds) *Comprehensive Biochemistry* New York Elsevier 1963 vols II III IV and VIII

FORRESTER T AND H SCHMIDT An electrophysiological investigation of the slow fibre system in the frog rectus abdominis muscle *J Physiol (London)* 207 477 491 1970

FRANKENHAEUSER B AND A F HODGKIN The after effects of impulses in the giant nerve fibres of *Loligo J Physiol (London)* 131 341 376 1956

GEREN B B The formation from the Schwann cell surface of myelin in the peripheral nerves of chick embryos *Exp Cell Res* 7 558 562 1954

GILLIS J M The site of action of calcium in producing contraction in striated muscle *J Physiol (London)* 200 849 864 1969

GOODMAN R (ed) *Genetic Disorders of Man* Boston Little Brown 1970

GORDON A M A F HUXLEY AND F J JULIAN Tension development in highly stretched vertebrate muscle fibers *J Physiol (London)* 184 143 169 1966

GORDON A M A F HUXLEY AND F J JULIAN The variation in isometric tension with sarcomere length in vertebrate muscle fibers *J Physiol (London)* 184 170 192 1966

GRANT J C B AND J V BASMAJIAN *Grant's Method of Anatomy*, Ed 7 Baltimore Williams & Wilkins 1965

GREENBERG, D M Amino acid metabolism *Ann Rev Biochem* 33 633 666 1964

GUTH, L , AND F J SAMAHA Qualitative differences between actomyosin ATPase of slow and fast mammalian muscle *Exp Neurol* 25 138 152, 1969

HARTSHORNE, D J M PHEINER AND H MUELLER Studies on troponin *Biochem Biophys Acta* 175 320 330, 1969

HAUROWITZ F Antibody formation *Physiol Rev* 45 1 17 1965

HENDLER R W Biological membrane ultrastructure *Physiol Rev* 51 66 97, 1971

HENNEMAN, E AND C B OLSON Relations between structure and function in the design of skeletal muscles *J Neurophysiol* 28 581 598 1965

HESS A The sarcoplasmic reticulum, the T system and the motor terminals of slow and twitch muscle fibers in the garter snake *J Cell Biol* 26 467 476 1965

HESS A AND G PILAR Slow fibers in the extraocular muscles of the cat *J Physiol (London)* 169 780 98, 1963

HILLE B Pharmacological modifications of the sodium channels of frog nerve *J Gen Physiol* 51 199-219, 1968

HODGKIN A L *The Conduction of the Nerve Impulse* Springfield Ill Thomas, 1964

HODGKIN, A L , A F HUXLEY, AND B KATZ Measurement of current voltage relations in the membrane of the giant axon of *Loligo J Physiol (London)* 116 424-448, 1952

HODGKIN, A L AND R D KEYNES The mobility and diffusion coefficient of potassium in giant axons from *Sepia J Physiol (London)* 119 513-528, 1953

HODGKIN A L , AND R D KEYNES Active transport of cations in giant axons from *Sepia* and *Loligo J Physiol (London)* 128 28 60, 1955

HODGKIN A L , AND R D KEYNES Experiments on the injection of substances into squid giant axons by means of a microsyringe *J Physiol (London)* 131 592-616, 1956

HOLTER, H Pinocytosis *Int Rev Cytol* 8 481-504, 1959

HOWELL J N A lesion of the transverse tubules of skeletal muscle *J Physiol (London)* 201 515-533, 1969

HUXLEY, A F Muscle structure and theories of contraction *Progr Biophys* 7 255-318, 1957

HUXLEY, A F , AND R STAMPFLI Evidence for saltatory conduction in peripheral myelinated nerve fibres *J Physiol (London)* 108 315-339, 1949

HUXLEY, A F , AND R STAMPFLI Direct determination of mem

brane resting potential and action potential in single myelinated nerve fibres *J Physiol (London)* 112 476–495 1951
HUXLEY A F AND R W STRAUB Local activation and interfibrillar structures in striated muscle *J Physiol (London)* 143 40p–41p 1958
HUXLEY A F AND R E TAYLOR Local activation of striated muscle fibers *J Physiol (London)* 144 426–441 1958
KAPLAN D M AND R S CRIDDLE Membrane structural proteins *Physiol Rev* 51 249–273 1971
KATZ B *Nerve Muscle and Synapse* New York McGraw Hill 1966
KATZ B *The Release of Neural Transmitter Substances* Springfield Ill Thomas 1969
KATZ A M Contractile proteins of the heart *Physiol Rev* 50 63–158 1970
KATZ A M AND A J BRADY Mechanical and biochemical correlates of cardiac contraction (I) *Mod Concepts Cardiovas Dis* 40 39–43 1971
KAUZMANN W Some factors in the interpretation of protein denaturation *Advances Protein Chem* 14 1–63 1959
KEYNES R D The ionic movements during nervous activity *J Physiol (London)* 114 119–150, 1951
KUFFLER S W AND E M VAUGHAN WILLIAMS Small nerve junctional potentials The distribution of small motor nerves to frog skeletal muscle and the membrane characteristics of the fibers they innervate *J Physiol (London)* 121 289–317 1953
KUFFLER S W AND E M VAUGHAN WILLIAMS Properties of the slow" skeletal muscle fiber of the frog *J Physiol (London)* 121 318–340 1953
KUSHMERICK M J R E LARSON AND R E DAVIES The chemical energetics of muscle contraction I Activation heat, heat of shortening and ATP utilization for activation relaxation processes *Proc Roy Soc (London) [Biol] Ser B* 174 293–313 1969
KUSHNER D Self assembly of biological structures *Bact Rev* 33 302–345 1969
LAJTHA A (ed) *Handbook of Neurochemistry* New York Plenum 1969 and 1970 vols I–VII
LARNER, J Inborn errors of metabolism *Ann Rev Biochem* 31 569–588, 1962
LEACH S J (ed) *Physical Principles and Techniques of Protein Chemistry* New York Academic Press 1969 vol 7 part A
LEHNINGER A L *The Mitochondrion* New York Benjamin 1964
LEHNINGER A L *Biochemistry* New York Worth 1970
LIMA DE FARIA A (ed) *Handbook of Molecular Cytology* Amsterdam Holland North Holland 1969
LOEWY, A G AND P SIEKEVITZ *Cell Structure and Function*, Ed 2 New York Holt, Rinehart Winston 1969
MARGOULIES M (ed) *Protein and Polypeptide Hormones* Amsterdam Holland Excerpta Medica Foundation 1968
McLENNAN H *Synaptic Transmission*, Ed 2 Philadelphia Saunders 1970
McPHEDRAN A M R B WUERKER AND E HENNEMAN Properties of motor units in a homogeneous red muscle (soleus) of the cat *J Neurophysiol* 28 71–84 1965
MEISTER A (ed) *Biochemistry of Amino Acids* New York Academic Press 1965 vols I and II
MOMMAERTS W F Energetics of muscular contraction *Physiol Rev* 49 427–508 1969
MUNRO H AND J B ALLISON (eds) *Mammalian Protein Metabolism* New York Academic Press 1969 vol III
NASTUK W L AND A L HODGKIN The electrical activity in single muscle fibers *J Cell Physiol* 35 39–74 1950
NERST W Zur Theorie des elektrischen Reizes *Pflueger Arch* 122 275–314 1908
NEURATH H (ed) *The Proteins* New York Academic Press 1963, vols I and II
NEW YORK HEART ASSOCIATION The contractile process Proceedings of a Symposium *J Gen Physiol* 50 (Suppl 6) 5–292, 1967
NORRIS F H , JR AND R L IRWIN Motor unit area in a rat muscle *Amer J Physiol* 200 944–946 1961
NOVIKOFF A B AND E HOLTZMAN *Cells and Organelles* New York Holt Rinehart Winston 1970
PAGE S G Comparison of the fine structures of frog slow and twitch muscle fibers *J Cell Biol* 26 477–497 1965
PARSONS C AND K R PORTER Muscle relaxation Evidence for an intrafibrillar restoring force in vertebrate striated muscle *Science* 153 426–427 1966
PEACHEY L D AND A F HUXLEY Local activation and structure of slow striated muscle fibers of the frog *Fed Proc* 19 257 1960
PETERS A *The Fine Structure of the Nervous System* New York Hoeber 1970
PHILPOTT C W AND M A GOLDSTEIN Sarcoplasmic reticulum of striated muscle Localization of potential calcium binding sites *Science* 155 1019–1021 1967
POLAND D AND H A SCHERAGA *Theory of Helix Coil Transitions in Biopolymers* New York Academic Press 1970
PORTER K R The sarcoplasmic reticulum Its recent history and present status *J Biophys Biochem Cytol* 10 (Suppl 4) 211–226 1961
PORTER K R AND G E PALADE Studies on the endoplasmic reticulum III Its form and distribution in striated muscle cells *J Biophys Biochem Cytol* 3 269–300 1957
POST R L C R MERRITT C R KINSOLVING AND C D ALBRIGHT Membrane adenosine triphosphatase as a participant in the active transport of sodium and potassium in the human erythrocyte *J Biol Chem* 235 1796–1802 1960
PREWITT M A AND B SAFALSKY Enzymic and histochemical changes in fast and slow muscles after cross innervation *Amer J Physiol* 218 69–74 1970
REIS D J AND G F WOOTEN The relationship of blood flow to myoglobin capillary density and twitch characteristics in red and white skeletal muscle in cat *J Physiol (London)* 210 121–135 1970
REITHEL F The dissociation and association of protein structures *Advances Protein Chem* 18 123–226 1963
RICHTER D *Metabolism of the Nervous System* New York Pergamon 1957
ROBERTSON J D New observations on the ultrastructure of the membranes of frog peripheral nerve fibers *J Biophys Biochem Cytol* 3 1043–1047 1957
ROMER A S The muscular system Ch 9 In *The Vertebrate Body*, Ed 3 Philadelphia Saunders 1962 chap 9
ROSE W C The significance of the amino acids in nutrition *Harvey Lect* 30 49–65 1935
ROSE W C The nutritive significance of the amino acids *Physiol Rev* 18 109–136 1938
SANDAW A S R TAYLOR, A ISAACSON AND J J SEGUIN Electromechanical coupling in potentiation of muscular contraction *Science* 143 577–579 1964
SCHOENHEIMER R AND D RITTENBERG Study of intermediary metabolism of animals with aid of isotopes *Physiol Rev* 20 218–248 1940
SCHROEDER E AND K LUEBKE *The Peptides* New York Academic Press 1966 vol II
SCHULTZE H E AND J F HEREMANS *Molecular Biology of Human Proteins* New York Elsevier 1966, vol I
SIEGEL G J , AND R W ALBERS Nucleosidotriphosphate phosphohydrolases In *Handbook of Neurochemistry*, edited by A Lajtha New York Plenum 1970, chap 2 vol 4
SKOU J C The influence of some cations on an adenosine triphosphatase from peripheral nerve *Biochim Biophys Acta* 23 394–401 1957
STEIN W D *The Movement of Molecules Across Cell Membranes* New York Academic Press 1967
STEIN J M , AND H A PADYKULA Histochemical classification of individual skeletal muscle fibers of the rat *Amer J Anat* 110 103–123, 1962
STEINHARDT J AND J A REYNOLDS *Multiple Equilibria in Proteins* New York Academic Press, 1970
TANFORD, C *The Physical Chemistry of Macromolecules* New York Wiley 1961
TOWER D B Molecular Transport Across Neural and Non neural Membranes In *Properties of Membranes* Berlin Springer 1962
VAN HARREVELD A On the force and size of motor units in the rabbit sartorius muscle *Amer J Physiol* 151 96–106 1947
VRBOVÁ G The effect of motoneurone activity on the speed of contraction of striated muscle *J Physiol (London)* 169 513–526 1963

WILKIE, D. R. Facts and theories about muscle. *Prog. Biophys.* 4: 288–324, 1954.

WILKIE, D. R. Thermodynamics and the interpretation of biological heat measurements. *Prog. Biophys.* 10: 259–298, 1960.

WINEGRAD, S. Autoradiographic studies of intracellular calcium in frog skeletal muscle. *J. Gen. Physiol.* 48: 455–479, 1965.

WINEGRAD, S. The location of muscle calcium with respect to the myofibrils. *J. Gen. Physiol.* 48: 997–1002, 1965.

WUERKER, R. B., A. M. McPHEDRAN, AND E. HENNEMAN. Properties of motor units in a heterogeneous pale muscle (M. gastrocnemius) of cat. *J. Neurophysiol.* 28: 85–99, 1965.

WYSSBROD, H. R., W. N. SCOTT, W. A. BRODSKY, AND I. L. SCHWARTZ. Carrier-mediated transport processes. In: *Handbook of Neurochemistry*, edited by A. Lajtha. New York: Plenum, 1971, vol. 5.

YOUNG, J. Z. The functional repair of nervous tissue. *Physiol. Rev.* 22: 318–374, 1942.

2

Digestion

2 / CONTROL OF FOOD AND WATER INTAKE

The energy requirements of the body are met by the food that is consumed and absorbed If an excess of food were consumed continually and there were no change in the energy expenditure there would be a consistent weight gain A shortage in the energy value of the food assimilated would result in a steady weight loss The fact that the average adult maintains a constant weight or gains or loses weight very slowly is sufficient evidence that a balance is usually maintained between the energy expended by the body and the energy content of the food that is assimilated The precision of this adjustment can be appreciated when we realize that the addition of 500 mg of dry food (less than one soda cracker) to each meal would cause an increase in weight over a 10-year period of more than 10 lb , if there were no compensatory increase in energy expenditure Of course some people do gain weight but it is believed that this is usually the result of some abnormality affecting the control of food intake

The balance between food intake and energy output is maintained by control of both the food intake and the output of energy There is an immediate effect of the excess food through its specific dynamic action, which results in an increase in energy output This, however, is rarely sufficient to account for all the energy contained in the excess food eaten If it were, there would never be a gain in weight If an individual consumes a certain measured excess of food at each meal over a period of time, he will gain weight steadily but the gain in weight will result in an increase in his basal metabolic rate which, if the excess food intake is not too great, will ultimately increase the metabolism to a point where the energy output will again equal the energy of the food, including the excess In this way a new balance is established at a higher weight level In addition, there is apparently another mechanism, less well understood, which, in some cases at least, balances the energy output against the food intake, even when this is in excess In experiments on dogs at the University of Illinois, it was found that dogs which were maintaining a constant weight on the amount of food which they ate voluntarily could be given an excess over this amount, through a gastric fistula, amounting to 33% of their voluntary food intake, without either diminishing their voluntary food intake or causing a gain in weight (Share et al , 1952) Evidently the animals had some means of metabolizing the excess food without increasing their total body mass This suggests that there is some mechanism for adjusting the total body metabolism to the energy content of the food consumed, even though it is in excess of the need, however, we are concerned here with the other aspect of the adjustment of food consumption to energy output, that is, the regulation of food intake

There are several methods to demonstrate the fact that food intake is controlled in some manner so as to equal the energy requirements of the animal Healthy animals, when allowed all the food which they wish to eat, will consume only as much as is necessary to meet their energy requirements and to maintain their weight, or their normal growth if they have not already attained their adult weight If the caloric value of the food is increased, the amount consumed will be diminished so as to maintain a normal intake of food calories If, on the other hand, the caloric value of the food is decreased by dilution with inert material, the animal will usually consume more food, up to the limit of its capacity, in order to maintain an adequate caloric intake This mechanism may be tricked by human devices into permitting an excess food intake, as in the experiments of Share and his coworkers described above, but usually it works very well Adjustment of food intake to energy expenditure is illustrated in Figure 2 1

Hunger and Appetite

Since eating is a voluntary act, control of food intake must be accomplished through con-

scious sensations which indicate whether food is needed or if enough food has been consumed The sensations which cause food to be eaten are described as hunger and appetite These terms have never been precisely defined in the scientific sense, but in a general way we recognize that hunger is the sum of the sensations aroused by the physical need for food Appetite, on the other hand, is a psychic or emotional desire to eat, and may or may not be associated with the need for food Cessation of eating is also a voluntary act, and is induced by a conscious sensation called satiety when a sufficiency of food has been eaten The mechanisms of hunger and satiety are probably inborn and dependent upon the inherent organization of the nervous system Appetite, on the other hand, is acquired, and is probably dependent upon pleasurable past experiences associated with eating For example, an individual who is hungry will eat almost any wholesome food so long as it is reasonably palatable When he has reached a state of satiety, however, he will eat only those foods which are particularly pleasing to him and with which he has had pleasant experiences in the past Thus, because of hunger, one may eat a wholesome meal which is fully adequate for his needs and then, because of appetite, add to it a dessert which is entirely unnecessary so far as caloric requirements are concerned

Neural Control of Food Intake

CENTRAL MECHANISMS

It has long been known that persons suffering from tumors of the hypophysis have a tendency to become obese, the condition is known as Frohlich's syndrome In 1940, Hetherington and Ranson showed that a similar condition could be produced in animals by injury to the hypothalamus in the vicinity of the hypophysis Later, Brobeck and others (1943) showed that the obesity was caused by injury to ventromedial hypothalamic nucleus situated in the tuber cinerium If this area is destroyed in rats, cats, or monkeys, instead of being satisfied with an amount of food appropriate to its caloric requirements, the animal continues to eat as long as it is able to swallow food or as long as food is available, with the result that it becomes excessively obese A little later, Anand and Brobeck (1951) found that injuries to the lateral hypothalamic nuclei, just lateral to those involved in obesity, would cause the animal to stop eating

Voluntary Caloric Intake and Body Weight as Functions of Exercise in Normal Animals

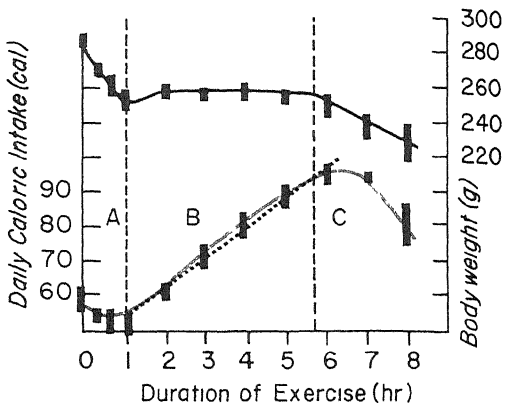


Fig 2 1 The effect of exercise of various durations on caloric intake and body weight of exercised rats Three ranges can be distinguished *A* sedentary *B* proportional response (normal activity) *C* exhaustion (From Mayer (1955) Courtesy of New York Academy of Science)

altogether Animals with such injuries absolutely refused food even though they starved to death If an animal with a medial hypothalamic injury and consequent hyperphagia (habitual overeating) was subjected to an injury to the lateral hypothalamus, it not only recovered from its hyperphagia, but refused absolutely to eat anything From these observations, it was concluded that the nuclei in the lateral area of the hypothalamus contain a "feeding center," and that the medial hypothalamic nuclei contain a "satiety center " Since injury to the medial nuclei was effective only if the lateral nuclei were intact, it was concluded that the medial satiety center acted by inhibiting the lateral feeding center

It is interesting to note that the areas of the hypothalamus involved in the control of food intake have a much higher concentration of high energy phosphate compounds (adenosine triphosphate, creatine phosphate) in fasted animals than in animals recently fed, the increase in these compounds implies increased activity of these areas in hungry animals (Larsson, 1954) The medial hypothalamic areas, which are believed to inhibit the feeding centers in the lateral hypothalamus, exhibit increased electrical activity when drugs which reduce or abolish appetite, such as amphetamine, are given It is believed that the increased electrical activity indicates stimulation of the medial nuclei and that these then inhibit the action of the feeding centers It has been shown that electrical stimu-

lation of the lateral hypothalamic nuclei by means of implanted electrodes produces hyperphagia in rats, cats, mice, monkeys, goats, and sheep During the period of stimulation, the animals exhibit a compulsive desire to eat, and if food is available will eat as much as it is possible for them to swallow (Larsson, 1954)

The neural organization of the feeding centers in the rat have been investigated by Morgane (1961) He devised a clever experiment which required the animal to cross an electrical grill and press a lever to obtain food After electrodes were implanted in various parts of the hypothalamus, stimulation of the far-lateral hypothalamus not only resulted in a feeding response in sated animals, but also motivated them to cross the electrical grill to obtain food When the mid-lateral hypothalamus was stimulated, a feeding response was obtained, but the animals were not "motivated" to cross the electrical grill Lesions were then placed in the medial fore-brain bundle, and it was found that a feeding response could still be obtained in sated animals by far-lateral hypothalamic stimulation but the animals would not cross the grill to obtain food These results suggest that the medial fore-brain bundle is important in "hunger motivation," since overcoming a barrier to obtain food apparently depends upon the integrity of this bundle When the medial fore-brain bundle was interrupted, basic feeding responses could still be obtained, but the animal would not "work" for his food Morgane concluded that the feeding center in the rat probably is composed of both a basic feeding drive (mid-lateral hypothalamus) and a hunger drive (far-lateral hypothalamus) element, and that only the latter is depressed by the satiety mechanism

Grossman (1960) has used an ingenious method of stimulating areas of the hypothalamus of the rat He implanted a small double cannula into the lateral hypothalamus and was able to study the effects of stimulation of the "feeding area" by crystalline chemicals Adrenergic agents, such as norepinephrine and epinephrine, caused satiated rats to eat, whereas cholinergic agents, such as acetylcholine or carbachol, caused the animals to drink large quantities of water Thus, the feeding mechanism appears to be selectively activated by adrenergic stimulation, while the drinking mechanism responds selectively to cholinergic stimulation

Brobeck (1957) suggested that the areas of the hypothalamus which control food intake do so either by facilitating or inhibiting the reflex mechanisms involved in the acquisition and consumption of food, he called these feeding reflexes He classified these in two ways *1)* according to the sensory mechanism arousing the reflex, and *2)* according to the nature of the response The reflexes are listed in Table 2 1 Probably the list could be extended if the mechanisms involved were better understood The important point is that the hypothalamic centers do not control food intake directly, but do so by making more or less responsive to stimuli the lower brain centers which control the reflexes governing the acquisition, consumption, and assimilation of food

PERIPHERAL MECHANISMS

Consumption of food brings about consequences which induce satiety and stimulate the mechanisms that are inhibitory to the feeding reflexes, starvation, or the absence of food, brings about consequences which excite hunger and stimulate the feeding centers In spite of extensive study, and even more widespread speculation, the peripheral mechanisms which contribute to hunger and satiety are still not fully understood Among the possible operating devices we can recognize gastrointestinal mechanisms and metabolic mechanisms The gastrointestinal mechanisms are mainly of a physical nature whereas the metabolic mechanisms are chemical The physical mechanisms are presumed to act through stimulation of sensory nerve endings, whereas the chemical mechanisms are supposed to act on some hypothetical sensory device, either within the central nervous system or connected with it through afferent nerves, although this is by no means certain

Among the gastrointestinal stimuli which tend to increase food intake and to cause a sen-

TABLE 2 1

Classification of feeding reflexes on a basis of initial sensory stimulus (left side) or behavior pattern induced (right side)

Sensory Basis	Behavior Basis
Visual reflexes	Reflexes of attention
Olfactory reflexes	Reflexes of approach
Auditory reflexes	Reflexes of examination
Tactile reflexes	Reflexes of incorporation
Gustatory reflexes	Reflexes of rejection
Enteroceptive reflexes	

Brobeck, 1957

sation of hunger are the so-called hunger contractions of the stomach and intestine (Quigley, 1955) The fact that hunger contractions cause a sensation of hunger helps us little in the solution of our problem, for we still do not know what causes the hunger contractions Besides mere emptiness of the stomach and intestine they are probably excited by something in the metabolic category Emptiness is not the whole answer because, although the stomach may be completely empty a few hours after a meal hunger contractions become more intense and more distressing the longer the fast is continued However, they cease at once when food is taken and before any metabolic changes have taken place, therefore, whatever their cause, they can be completely inhibited by the presence of food in the stomach and intestine

It is known that hunger contractions can be induced, or made more intense if already present, by inducing a state of hypoglycemia through administration of insulin (Quigley et al , 1929, 1930) This brings about stimulation of medullary centers, in which increased activity of the vagus centers is a prominent feature If the vagus nerves are cut, insulin hypoglycemia does not stimulate hunger contractions (Quigley and Templeton, 1930), indeed it tends to inhibit them if they are already present The hunger contractions induced by insulin evidently are due directly to an action of low blood sugar on the central nervous system They can be inhibited by administration of glucose intravenously, but the normally occurring changes in blood sugar level are in no way related to the presence or absence of gastric or intestinal hunger contractions (Quigley, 1955, Scott et al , 1938) Hyperglycemia may or may not inhibit hunger contractions, it obviously does not in diabetic subjects In a majority of experiments on normal animals, hyperglycemia induced by injecting glucose intravenously does not inhibit gastric hunger motility It has been shown by Stunkard and Wolff (1956) that, in those instances in which the hunger contractions were not inhibited by intravenous glucose, the injection of glucose did not cause a great rise in the arterial glucose concentration, as compared to the venous glucose concentration The arteriovenous glucose concentration difference is taken as an index of the rate of utilization of glucose by the tissues On the other hand, in those instances in which there was inhibition of hunger contractions, there was an increase in arteriovenous glucose difference Reference will be made to this fact a little later under the heading of "Metabolic Factors "

Undoubtedly gastric hunger contractions contribute to the hunger state but they are by no means solely responsible for it Enough human cases have been studied in which both splanchnic nerves were severed to establish the fact that this operation abolishes all sensation produced by contraction of the stomach during hunger The contractions persist but they are no longer felt However individuals so operated upon do not experience any decrease in their desire for food and they have periods of hunger just as intense as before the splanchnic nerves were severed Cutting the vagus nerves abolishes the hunger contractions altogether for a time, but by no means abolishes the sensation of hunger We may conclude that the hunger sensation arises not only in the stomach and intestine but elsewhere in the body as well, and that the gastrointestinal component is dispensable

Sensations arising in the gastrointestinal tract which diminish or abolish food intake are somewhat better understood The act of eating will, in itself, diminish hunger and appetite For example, if an animal is prepared with an esophagostomy so that food which is eaten passes through the mouth and pharynx and upper esophagus but does not enter the stomach, the animal will eat more food than would a normal animal, but feeding ceases eventually If an animal is given a measured amount of food before its regular mealtime it will reduce the amount of its voluntary food intake proportionately If the prefeeding is carried out through a gastric fistula instead of by mouth the reduction in voluntary food intake is much less pronounced, indicating that the decrease in food intake was not all due to distention of the stomach but partly to the oropharyngeal stimulation of eating

Distention of the stomach alone inhibits hunger and appetite For example, if an animal is given a portion of its food a few minutes before regular feeding time through a gastric fistula, and this practice is continued over a period of time, eventually the animal will come to reduce the size of its subsequent meal by an amount equal to the amount given through the gastric fistula; in other words, the total size of the meal will be maintained constant (Janowitz and Hollander, 1955) This adjustment does not occur immediately and is never very precise, but it does show that gastric distention

with food tends to reduce hunger and appetite Similarly, if the stomach is partly filled with inert materials, the size of the subsequent meal will be reduced Inflation of a balloon in the stomach has the same effect (Janowitz and Grossman, 1949) That these effects are mechanical and not metabolic in nature is shown by the fact that, if the artificial feeding, instead of being given immediately before the regular meal, is given some 4 hr after, and time is allowed for it to be disposed of, the animal will eat a normal amount at the next regular meal, provided the artificial feeding was not too large As much as one-third of the animal's regular food consumption can be given in this manner over a period of time without affecting his voluntary intake (Share et al , 1952) The fact has already been referred to that animals so fed, although receiving 30% or more in excess of their regular food intake, do not necessarily gain weight The effect of gastric distention on hunger and appetite is doubtless due to the presence in the stomach of stretch receptors (Paintal, 1954) situated in the muscular walls and connected to afferent fibers of the vagus nerves, section of the vagi does away with the effects of gastric distention on appetite Indeed, vagotomy may produce a condition of hyperphagia comparable to that caused by lesions of the medial hypothalamic nuclei

The effects of the gastrointestinal stimuli are largely confined to determining the size of an individual meal They do not serve to regulate the overall balance between the energy intake represented by the food and the energy output represented by the total metabolism of the body Presumably, metabolic factors are responsible for this type of regulation, but any discussion of such factors at the present time would be speculation We know certainly only that some metabolic consequences of the assimilation of food are not effective as regulatory factors It has been shown, for example, that changes in blood sugar, at least those that are not associated with changes in arteriovenous glucose difference, have no effect on voluntary food intake, nor do they influence hunger contractions—except those induced by insulin (Quigley and Hallaran, 1932) An increase in the amino acid concentration of the blood caused by intravenous injection of amino acids has no effect upon voluntary food intake in dogs However, the slight increase in temperature which results from the presence of pyrogens in some intravenous solutions does definitely influence appetite, even though the rise in temperature is too small to be recognized clinically as a fever This, and the fact that taking food causes a slight increase in temperature due to the specific dynamic action of the food, has suggested to Brobeck (1957) the possibility that the increase in body temperature caused by the specific dynamic action of the food may be a factor which limits food intake (Fig 2 2)

Mayer (1955) has suggested that food intake is controlled in some way through the metabolism of glucose—the so-called glucostatic hypothesis Recognizing that hunger and appetite are not related to changes in blood sugar level, Mayer contends that they are governed by the utilization of carbohydrate food by the tissues "According to the glucostatic theory the mechanism postulated for the short term regulation of energy intake rests on the concept that somewhere, possibly in the hypothalamic centers shown to be implicated in the regulation of food intake, perhaps peripherally as well, there are glucoreceptors sensitive to blood glucose in the measure that they can utilize it " In other words, when the tissues are utilizing glucose rapidly, thus establishing a high arteriovenous glucose difference, the glucoreceptors are stimulated, and through their central connections suppress hunger and appetite and inhibit food intake On the other hand, when glucose is not available for tissue utilization, and the arteriovenous glucose difference tends toward zero, a state corresponding to tissue hunger exists, this is sensed in some way by the hypothalamic centers, which then facilitate the feeding reflexes

Another theory suggests that the total energy stores of the body are maintained at a certain predetermined level by the balance between

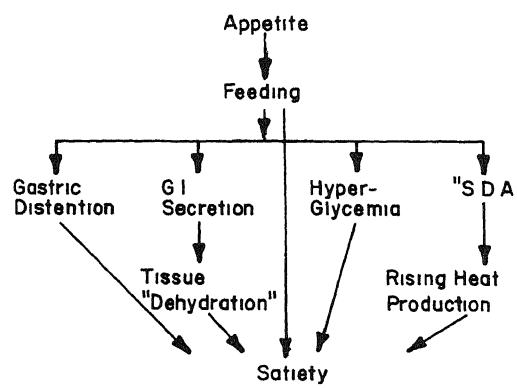


Fig 2 2 Outline of mechanisms taking part in transition from appetite to satiety (From Brobeck, 1957)

hunger and satiety Whatever the mechanism may be, it is evident that metabolic factors are important and quite possibly play a determining role in the regulation of the balance between food intake and energy output in the body

Thirst

The sensation of thirst is referred to the pharynx and is due to the stimulation of sensory nerve endings in this situation Two theories have been advanced to explain the mechanism by which the sensation is aroused

According to one view, thirst is due to the drying of the pharyngeal mucous membrane, the salivary glands being given a role in the regulation of the water balance of the body When the water content of the body falls below a certain level, salivary secretion is depressed The consequent drying of the mucous membrane of the throat then elicits the characteristic sensation If such a view is correct, drying of the pharyngeal mucosa, from whatever cause, should cause thirst Cannon (1929), who provided the principal support for this theory, found, in studies upon himself, that after abstinence from fluids for a time the depression of salivary secretion which resulted was definitely associated with thirst Atropine, which inhibits salivary secretion, also produced the typical sensation, and thirst aroused by the deprivation of water was relieved by the application of cocaine to the mucosa Pilocarpine, or acid substances which stimulate the flow of saliva, also relieve thirst In dogs, however, atropine and pilocarpine have no effect upon the water intake

According to the other view, thirst is a sensation resulting from changes in blood composition, probably due to a rise in its osmotic pressure, which stimulates the afferent nerve endings or perhaps acts upon central nervous structures Thus, Rowntree and his associates (Weir et al , 1922), found that thirst of diabetes insipidus was relieved neither by cocainization of the mucous membrane nor when salivation was induced by pilocarpine The experiments of Gilman (1937) suggest that cellular dehydration, rather than a rise in osmotic pressure, is the prime factor in arousing thirst Elevation of the osmotic pressure of the blood of dogs by the injection of a hypertonic salt solution, and the imbalance thus caused in the osmotic relationship between intra- and extracellular fluids, caused a much greater intake of water than an equivalent rise in osmotic pressure resulting from the administration of urea After the injection of salt, sufficient water was drunk to reduce quickly the osmotic pressure of the blood After injection of urea, to which the cells are readily permeable, and which did not disturb the osmotic relationships the water intake was but little increased and the osmotic pressure of the blood remained elevated In support of the view that a lowered water content of the cells is the true thirst stimulus, Gilman cited an experiment in which anhydremia was induced by the withdrawal of large quantities of extracellular electrolytes (e g , NaCl) without the withdrawal of water In such anhydremic animals, dehydration of the tissue cells does not occur, and although the oral mucous membranes are quite dry, there is no evidence of thirst and water is refused

The two theories are not incompatible, as is indicated by the work of Holmes and Gregerson (1947), who found that, in human subjects, following the intravenous injection of 300 ml of 5% sodium chloride, although there was an increase in plasma volume there was a considerable decrease in salivary secretion associated with the sensation of thirst This proves that the decrease in salivary secretion leading to a sensation of thirst may be caused by an increase in the osmotic pressure of the blood, and suggests that tissue dehydration may induce the sensation of thirst through the reduction in the flow of saliva

The neural mechanisms for the regulation of water intake are similar to those for the regulation of food intake Indeed, they are closely associated and have not as yet been clearly separated Certain types of injuries to the lateral hypothalamus induce not only a cessation of food intake but also cessation of water intake The animals die of dehydration unless given water by stomach tube (Greer, 1955, Morrison and Mayer, 1957) Greer found that stimulation of the dorsomedial nucleus in the cat caused violent drinking activity, thus there are probably centers in the hypothalamus for the regulation of water intake comparable to the feeding centers Like the intake of food, the intake of water is influenced by gastric distention Animals made thirsty by intravenous injection of hypertonic salt solution will usually drink an amount of water over a short period of time approximately equal to the amount necessary to dilute the injected salt to an isotonic concentra-

tion. If the stomach is distended by means of a water-filled balloon, the amount of water that is voluntarily drunk will be diminished and the replacement of the water deficit delayed. The vagi play a part in the regulation of water intake. Vagotomized animals not only eat more food but also drink more water than do normal animals. The food intake and the water intake are increased proportionately by vagotomy; they are decreased by sympathectomy (Towbin, 1955).

Excessive thirst also occurs in lesions of the anterior hypophysis, leading to diabetes insipidus. In this disease there is failure of production of the antidiuretic hormone of the anterior hypophysis, resulting in an excessive loss of water through the kidneys. In this case, thirst is probably secondary to dehydration.

2 / INNERVATION AND VISCERAL SENSATIONS OF THE GASTROINTESTINAL TRACT

The gastrointestinal tract receives an abundant nerve supply via the autonomic nerves In addition, it contains within its walls an elaborate plexus of interconnected ganglia This is usually described as consisting of two plexuses, the *myenteric plexus*, situated between the circular and longitudinal layers of the smooth muscle, and the *submucous plexus*, situated in the submucosa

Extrinsic Nerves

The autonomic nerves are classified as sympathetic and parasympathetic but it is more in keeping with modern thought to consider them as cholinergic or adrenergic, depending upon whether they liberate acetylcholine or an adrenaline-like substance at their final terminations In this sense, most parasympathetics are cholinergic and sympathetics are adrenergic, but there is considerable mixing, since cholinergic fibers have been demonstrated to be in considerable abundance in the sympathetics

The parasympathetic nerve supply to the gastrointestinal smooth muscle comes mainly from the vagus (Fig 2 3), however, the pelvic nerves supply the distal portion of the colon (Fig 2 4) The parasympathetic fibers, whether vagus or pelvic in origin, do not end directly on the smooth muscle cells but end by making synapses with cells in the enteric plexuses Therefore, as pointed out by Carlson et al (1922), they stand in relation to these plexuses as intercalated or internuncial fibers between reflex centers and not as ordinary motor nerve fibers The function of such fibers in the central nervous system, and presumably in this situation also, is to increase or decrease the excitability of the reflex centers and not to initiate muscular activity directly

The sympathetic supply to the stomach and small intestine is by way of the splanchnic nerves. The cecum, appendix, and ascending and transverse colon are supplied by nerves which arise from the superior mesenteric plexus This plexus is a derivative of the aortic plexus through which it also receives vagus fibers from the celiac plexus The lower part of the rectum is supplied by sympathetic fibers which arise from the upper and lower divisions of the hypogastric plexus The preganglionic sympathetic fibers end in the celiac, superior or inferior mesenteric, or aortic ganglia The fibers that reach the intestine, therefore, are postganglionic They probably have no functional relation to the enteric plexuses, even though they traverse the enteric ganglia and are distributed along with the postganglionic enteric plexus fibers Most of them are concerned with the vasomotor supply of the numerous blood vessels in the intestine but some go to the intestinal smooth muscle in the muscularis mucosae (Kuntz, 1953)

FUNCTION OF EXTRINSIC NERVES

The gastrointestinal tract is capable of carrying on its major functions after all the extrinsic nerves have been severed This automaticity may be attributed partly to the local nervous mechanism and partly to the properties of the smooth muscle Probably, the essentially rhythmic functions, such as gastric antral peristalsis and segmentic contractions of the small and large intestine, are dependent upon the properties of the smooth muscle, whereas the more highly coordinated functions, such as forward peristalsis in the small intestine and mass movements in the colon, are dependent upon the functional integrity of the myenteric plexus Both the neurogenic and myogenic functions are regulated through central nervous system reflexes by way of the autonomic nerves.

The tone and peristaltic activity of the stomach are usually increased on stimulation of the vagus nerves The statement that the vagus

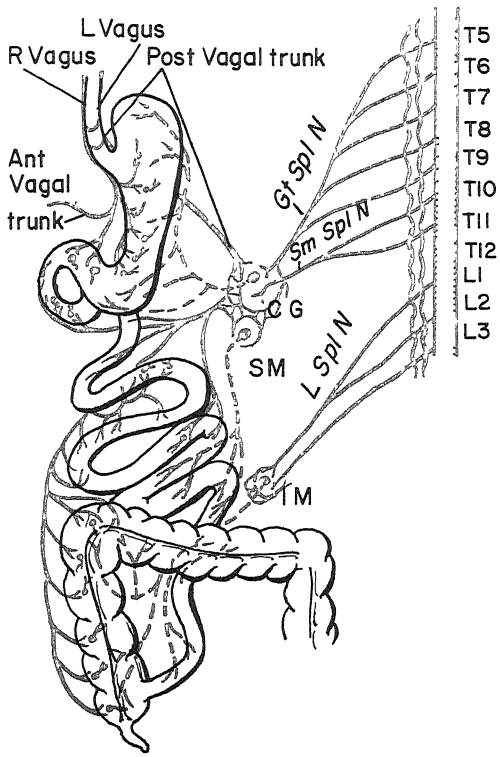


Fig 2 3 Diagram of the innervation of the stomach small intestine and proximal part of the colon GT SPL N great splanchnic nerve SM SPL N small splanchnic nerve L SPL least splanchnic nerve C G celiac ganglion S M superior mesenteric ganglion I M inferior mesenteric ganglion Continuous lines vagal and sympathetic preganglionic fibers broken lines sympathetic postganglionic fibers ganglion cells and postganglionic fibers of vagus in gastrointestinal wall

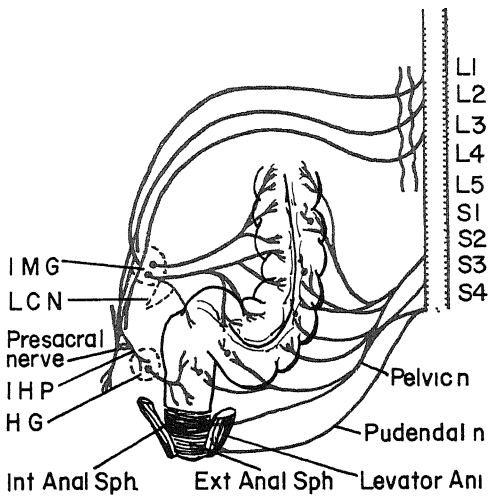


Fig 2 4 Diagram of the innervation of the distal colon

increases the activity of the muscle in the gastric wall while inhibiting the sphincters has been proved to be erroneous, when motor effects are observed in the muscle of the pyloric portion, for example, the pyloric sphincter responds in the same sense, thus the idea of a reciprocal innervation of the sphincters and the other muscle of the stomach is not borne out by the experimental facts

The effect on the stomach of stimulating the splanchnic nerves is usually described as inhibitory but it is far from being invariably so In simple acute experiments on anesthetized animals, increased motor activity is seen about as frequently as inhibition when the splanchnic nerves are stimulated Carlson and his associates (1922) pointed out that motor responses are more likely to be obtained when the muscle is in a state of relaxation, whereas inhibitory effects occur when the muscle is in a state of high tonus or vigorous activity This rule applies to both vagus and sympathetic stimulation, nevertheless, with the muscle in a medium state of tonus, the vagi will more often produce excitatory effects while the splanchnics produce inhibitory effects The reason for the preponderance of sympathetic motor effects in recently anesthetized, operated animals is that the stomach, in such circumstances, is always in a state of very low tone

In the small intestine, the parasympathetic nerves are nearly always excitatory and the sympathetics nearly always inhibitory (Fig 2 5), however, the excitatory vagus effects, as mentioned previously, are usually preceded by a brief inhibition In the upper portion of the small intestine, motor effects from stimulation of the sympathetic nerves are by no means uncommon The tonus rule holds here also, motor effects are more apt to be observed in the relaxed intestine

In the large intestine, stimulation of the parasympathetic innervation is reported to cause mainly increased activity, whereas stimulation of the sympathetic nerves gives results that are even more confusing than those obtained in the stomach Some investigators report only inhibitory effects, others report only motor effects, whereas still others report inhibition of the walls of the viscera with contraction of the sphincters

Visceral Reflexes

While studying the effects of electrical stimulation of autonomic nerves, it is well to bear in

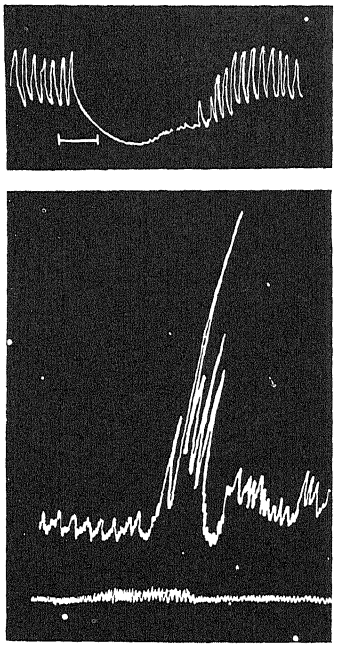


Fig 2 5 Upper inhibitory effect of splanchnic stimulation upon the movements of the small intestine the heavy white line indicates the time during which the stimulus was applied Lower effect upon intestinal motility of stimulating vagus nerves (After Thomas and Wheelon 1922)

mind the fact that nerves were not developed for the purpose of being artificially stimulated Under normal circumstances, the autonomic nerves serve only reflex functions, and when aroused normally through their reflex connections they may produce effects quite different from those seen on electrical stimulation Normal reflex stimulation is selective, exciting only those fibers which will promote the function subserved by the reflex, whereas ordinary electrical stimulation excites all the fibers without regard to their normal function

Certain gastrointestinal reflexes involving the peripheral autonomic nerves have already been mentioned, among these are the receptive relaxation of the stomach, the enterogastric reflex, and the so-called gastrocolic reflex These reflexes chiefly involve the parasympathetic nerves and are concerned with regulation of the normal functioning of the gastrointestinal smooth muscle Other reflexes have been described which occur only under unusual circumstances, and generally involve the functioning of the sympathetic nerve supply One of these is the *intestino-intestinal inhibitory reflex* This consists of reflex inhibition of the whole intestinal tract as a result of distention of some one segment The stomach may also be inhibited, but in this case we are dealing with a mechanism different from the enterogastric reflex previously mentioned The reflex is abolished or greatly diminished when the splanchnic nerves are cut, hence it involves discharge of impulses over the sympathetic nerves It probably belongs in the category of responses to noxious stimuli which, generally throughout the body, tend to cause increased activity of the sympathetic nervous system, and to inhibit gastrointestinal motility Other examples of reflex responses to noxious stimuli are inhibition of the intestine due to irritation of the peritonium, distention of the gall bladder and bile ducts, overfilling of the bladder, or forcible stretching of the anal sphincters All of these stimuli produce inhibition of the gastrointestinal smooth muscle through activation of adrenergic mechanisms, mainly the sympathetic nerves

Visceral Sensations

The abdominal and thoracic viscera are insensitive to the several types of stimuli which readily arouse sensations in the skin and more superficial tissues of the body The effects of the different varieties of stimulus, *thermal*, *chemical*, *tactile*, and *pain*, have been investigated by a large number of observers

TOUCH

The sensation of touch disappears at the lower end of the pharynx This was shown by Hurst (1911), who used an esophageal tube with a slit on one side through which the mucosa of the esophagus could be stimulated Carlson and Braafladt (1915), by means of a test-tube brush passed into the stomach, were unable to elicit the sensation of touch from the gastric mucosa The rectum possesses no tactile sensibility but the anal canal is sensitive

THERMAL SENSIBILITY

The esophagus is sensitive to extremes of heat and cold The sensations of temperature that are felt when hot or cold materials enter the stomach have been thought by some to originate in the lower end of the esophagus, by others to arise in the skin of the epigastrium, either through thermal conduction or by reflex changes in the cutaneous blood vessels. Carlson and his associates have shown, however, that the gastric mucosa is sensitive to extreme tem-

perature changes, i e , protophathic thermal sensibility (below 10° C or above 45° C) It is the lower end of the esophagus, however, which is responsible for the greater part of the thermal sensation that is experienced when excessively hot or cold materials are swallowed This is due to the greater sensitivity of the esophageal mucosa, as well as to the fact that the material may be retained for an appreciable length of time above the cardia The colon is insensitive to temperature changes, but even comparatively slight differences in temperature can be detected in the anal canal

CHEMICALS

Chemicals, with the exception of alcohol, cause no sensation whatever when introduced into the stomach or intestinal canal The mucosa is completely insensitive to acids, the introduction into the healthy stomach of a solution of 0 5% hydrochloric acid causes no pain or sensation of any kind Alcohol stimulates the mucosa of the esophagus and stomach and causes a sensation of warmth Peppermint and various condiments free from alcohol arouse no sensation The pelvic colon and rectum show a similar sensitivity to alcohol but are insensitive to other chemicals The anal canal is extremely sensitive, especially to alcohol and glycerine which cause a burning sensation

Pain

Sensations of pain cannot be elicited from the viscera by procedures that evoke pain in the skin The intestine or the liver, the stomach or the heart may be cut, burned, or pinched without arousing any immediate sensation In the second stage of a colostomy operation, for example, the colon can be opened without pain being experienced by the patient Harvey (1628, see 1962 ed) remarked upon the absence of sensation in the exposed heart of Viscount Montgomery Harvey records, "I carried the young man to the King (Charles I) that His Majesty might with his own eyes behold this wonderful case, that, in a man alive and well, he might, without detriment to the individual, observe the movement of the heart, and with his proper hand even touch the ventricles as they contracted And His Most Excellent Majesty, as well as myself, acknowledged that the heart was without the sense of touch, for the youth never knew when we touched his heart "

The insensibility of the alimentary tract to ordinary forms of stimulation commences in the middle third of the esophagus and extends as far as the commencement of the anal canal How can these observations be reconciled with the well-known fact that pain is one of the commonest manifestations of visceral disease?

The whole subject of pain arising in, or referred from, the viscera is still highly controversial No final answer can be given The most that can be attempted is a summary of some of the more significant experimental results and clinical observations, and the opinions of those who have especially interested themselves in the subject

REFERRED PAIN

When a viscus is diseased, pain or tenderness is frequently felt in the tissues approximately overlying it (abdominal or chest wall) or in some part quite remote from it For example, pain is felt in the neck or shoulder (Fig 2 6) in conditions affecting the diaphragm, between the scapulas in gastric disease, in the region of the umbilicus in appendicitis, in the testis in renal colic and in the sternal region, or down the left arm in angina pectoris Esophageal pain is frequently confused with that of cardiac origin, as it not only characteristically occurs substernally, but may extend into the neck, jaw, and even into the arms When traction is made upon a coronary artery of a dog, the animal whines and indicates the location of the pain by limping on the left forepaw In dextrocardia, anginal pain is felt in the right arm Other examples are the pain in the perineum and tip of the penis caused by a stone in the region of the neck of the bladder, and the pain in the groin due to a stone in the ureter

A given spinal segment supplies a visceral area with autonomic nerve fibers (afferent and

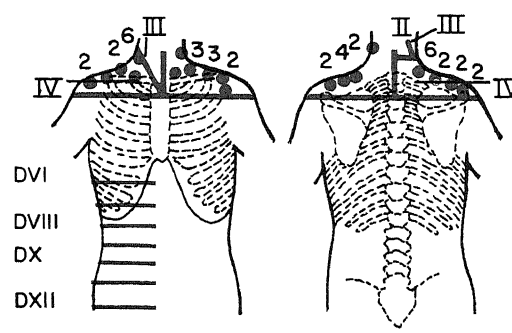


Fig 2 6 Location of maximum points of referred neck pain from diaphragm irritation in 23 cases They are all situated in the region supplied by the third and fourth spinal segments (After Capps and Coleman 1932)

efferent) and also a well-delineated area of the skin (dermatone) with somatic nerves The two types of structure linked in this way through the afferent nerves and the central nervous system may be some distance apart (e g , diaphragm and shoulder) or be more closely related (e g , area of abdominal wall and an underlying abdominal viscus) Mackenzie (1918) believed the afferent autonomic impulses arising in a diseased organ, although of themselves incapable of arousing any sensation, would, upon entering the spinal cord, set up an "irritable focus" with the result that cells accustomed to receive impulses from the corresponding somatic area were excited Thus, the impulses from the viscera spread, or "irradiated," on to cells of the corresponding somatic center New impulses originating in these cells traveled along the usual paths to higher perceptive centers (thalamus) which projected or referred the sensation to the somatic area, e g , skin or muscle, from which it was accustomed to receive impulses In this way, *spontaneous pain* in superficial structures remote from the diseased site was accounted for

Tenderness to touch, pressure, or light pinching of the skin (hyperesthesia and hyperalgesia) was ascribed to the impulses (which ordinarily would be below the threshold for pain) arriving in the segment rendered hyperexcitable as a result of impulses received from the diseased organ In the case of the abdominal viscera, Mackenzie claimed that the area of tenderness in the abdominal wall remained fixed, although the position of the diseased organ changed, thus indicating the referred nature of the pain He explained the rigidity (hypertonus) of muscles overlying a diseased organ (the right rectus abdominis in acute appendicitis, for example) upon a similar basis Afferent impulses of normal intensity arising in the muscle proprioceptors upon arriving at the spinal centers, which had been rendered hyperexcitable by the receipt of abnormal visceral impulses, resulted in a reflex increase in tonus of the corresponding muscles (see Fig 2 7)

Table 2 2 gives the segmental areas to which pain is referred in disease of various viscera

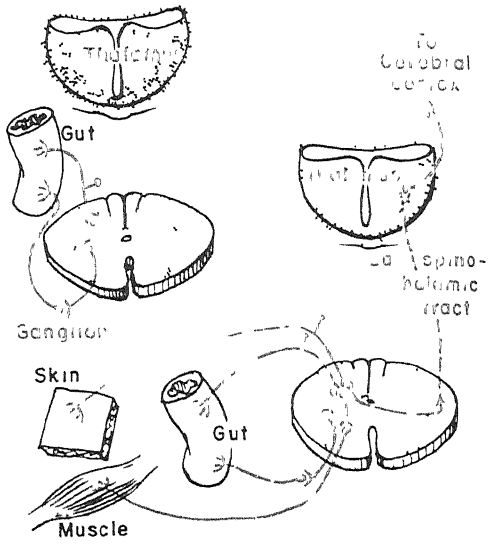


Fig 2 7 Diagram to illustrate Mackenzie's theory of referred pain (1920) A representing normal conditions a visceral reflex is shown B illustrates a visceromotor and a viscerosensory reflux Impulses from a viscus are conceived as setting up an irritable focus in the cord which through the involvement of neighboring neurons increases the tone of muscles innervated by the same segment and causes a discharge of impulses over the pathway for pain The sensation is then projected in consciousness to the periphery as indicated by the dotted line

VISCERAL PAIN

The pains of dysmenorrhea or childbirth, for example, or the pain of intestinal colic or an overdistended bladder, or the substernal pain of coronary occlusion, although diffuse, seem to arise within the organ itself, even when referred as well to a somatic structure, such as muscle or skin Kinsella (1948) observed that, in conscious patients, pressure with the fingers upon an inflamed appendix caused pain centered in the viscus, a healthy appendix is insensitive He and others have also shown that a chronic duodenal ulcer is also tender to direct contact In transsection of the cord at the third thoracic vertebra in a patient (reported upon by Kinsella, 1948), in whom there was complete paralysis and anesthesia of the abdominal wall, intestinal pain was experienced which obviously could not be caused by reference to somatic nerves Impulses in such cases must reach the central nervous system along intact visceral afferents They apparently travel by sympathetic fibers which enter the cord above the level of the transsection It is unlikely that they are transmitted by the vagus.

Pain in a normal viscus may also result from disease in a distant organ through a visceromotor reflex For example, painful contractions of the small intestine may accompany disease of the appendix A distinction should be drawn between pain of this nature and referred pain

Table 2 2
Spinal segments to which visceral pain is referred

Viscus	Spinal Segments
Lungs	1–7 dorsal, mostly 2–5 dorsal
Heart	3–5 cervical, 1–8 dorsal, predominantly on left side, sometimes bilateral
Esophagus	Mainly 5 dorsal, also 6, 7 and 8 dorsal
Breast	4 and 5 dorsal
Stomach	7, 8 and 9 dorsal, usually bilateral
Intestine	9–12 dorsal, bilateral or on left side only
Liver	8–10 dorsal on right side
Gall Bladder	Mostly 8 and 9 dorsal, also 5–7
Kidney	Mostly 10 dorsal, also 11 and 12 dorsal and 1 lumbar
Ureter	11 and 12 dorsal and 1 lumbar
Testis	10 dorsal
Epididymis	11 and 12 dorsal
Bladder	11 and 12 dorsal and 1 lumbar, also 3 and 4 sacral
Prostate	10 and 11 dorsal, also 1–3 and 5 sacral
Ovary	10 dorsal
Fallopian tubes	11 and 12 dorsal
Uterine cervix	11 and 12 dorsal and 1–4 sacral
Uterine body	10 dorsal to 1 lumbar

The Adequate Stimulus for True Visceral Pain. The existence of true visceral pain is not incompatible with the statement made above, that the viscera are insensitive to the ordinary types of stimulus From his investigations, Hurst (1911) concluded that the only adequate stimulus for visceral pain fibers is tension Distension of a hollow viscus, e g , stomach, intestine, gall bladder, etc , gives rise to pain, as a result of the stretch stimulus applied to the nerve terminals in its wall The pain is roughly localized to the viscus itself, or referred Under certain circumstances, chemical substances produced in an ischemic organ, as in angina pectoris or intermittent claudication, may stimulate nerve fibers subserving pain

Although the question is by no means settled, it appears that *contraction* of the muscular wall of a hollow viscus such as the intestine does not cause pain unless the movement of the bowel makes traction upon the mesentery Pain arises, however, if the contraction causes distention of a neighboring portion of the wall, as may result when the contraction wave approaches a mechanical obstruction, a length of bowel in spasm, or a sphincter which fails to relax (*achalasia*) Poulton (1928), for example, found that when a balloon was inserted into the lower part of the human esophagus, the approach of a peristaltic wave toward the obstruction caused pain, but during its passage over the esophageal wall in contact with the balloon, no sensation was felt Poulton ascribes the absence of pain during the passage of the contraction wave to the reduction in the diameter of the tube and the consequent release from stretch of the nerve endings lying between the muscle fibers Pain was also relieved if the esophageal muscle relaxed to accommodate the balloon, that is, adjusted the length of its fibers to the distending force

As further evidence for the effectiveness of distention in causing pain, the following observations may be cited In animals, when an intestinal loop exposed under local anesthesia is stimulated to powerful contraction, there is no evidence of pain, whereas even moderate distention of the loop (as by inflating it with a balloon) is manifestly painful Distention of the human appendix by the injection of fluids through an appendicostomy opening causes severe pain in the epigastrium or in the region of the umbilicus, and when the duodenum is distended by the injection of material through a duodenal tube, pain is felt on the right side The pain impulses undoubtedly reach the central nervous system through the splanchnic nerves Bentley and Smithwick (1940) distended the duodenum of patients whose splanchnics had been divided on one or both sides After unilateral denervation, distention caused pain on the unoperated side only, no pain whatever was felt after bilateral splanchnic section

Nausea

Nausea usually precedes the act of vomiting but may occur alone On the other hand, vomiting may occur without nausea, as in certain cerebral conditions The sensation is felt in the back of the throat or pit of the stomach, and in its milder degrees is merely a "sinking" sensation in the epigastrium It is frequently associated with vasomotor disturbances and sweating Increased tension upon the walls of the stomach or duodenum is a potent cause of the sensation, and Poulton (1928) has shown that it is also induced by distention of the lower part of the esophagus During the passage

of a peristaltic wave which relieves the tension upon the nerve fibers in the esophageal walls, the sensation is relieved. Barclay (1933) showed, by radioscopy in the human subject, that nauseous odors caused the lower border of the stomach to descend an inch or two, evidently as a result of sudden relaxation of the abdominal muscles. This movement would tend to stretch the esophagus and gastric walls and so exert tension upon the nerve endings. The stimulus which induces nausea, therefore, is apparently the same as that which causes visceral pain, but of lower intensity. It is likely that the sensations experienced during changes in speed of an elevator are also the result of tension exerted upon the esophagus and gastric walls. Tension is also probably a contributory factor in the production of seasickness, being brought into play by the pitch and roll of the ship.

The relief of nausea and vomiting by the application of counter-irritants to the epigastrium or over the sternum is probably due either to a reflex change in the tone of the gastric walls or to the reflex initiation of peristaltic contractions.

The tension upon the nerve endings in the latter instance is taken up by the muscle fibers. Poulton (1928) observed, for example, that the sensations caused by a balloon in the esophagus were relieved by vigorous friction of the skin over the sternum; contraction of the esophagus in some cases, or adjustment of the postural tone of the esophageal wall in others, were observed to accompany the disappearance of the sensation.

3 / GASTROINTESTINAL HORMONES

Some background information on the range of actions of gastrointestinal hormones is required for a better understanding of the sections which will follow on gastrointestinal secretion and motility

The great explosion of interest in the principal gastrointestinal hormones has followed from their chemical identification, purification, and, finally, synthesis These hormones are *gastrin* (Edkins, 1906), which "primarily" stimulates the gastric secretion of acid, *secretin* (Bayliss and Starling, 1902), which primarily stimulates the pancreatic secretion of bicarbonate and water, *cholecystokinin* (Ivy and Oldberg, 1928), which contracts the gallbladder, and *pancreozymin* (Harper and Raper, 1943), which stimulates the pancreatic secretion of enzymes

When these substances were available in pure form (Table 2 3), it became clear that pancreozymin and cholecystokinin, as had been suspected, are one substance Secondly, pancreatic glucagon, secreted by the alpha cells of the islets of Langerhans, was found to resemble the other gastrointestinal hormones, especially secretin, which it rather closely resembles in portions of its amino acid sequences Thirdly (Table 2 4), the terminal tetrapeptides of gastrin and cholecystokinin-pancreozymin are identical

However, the most striking finding was the fact that each of the three pure gastrointestinal hormones has an effect on each of the four primary or identifying actions by which the hormones were first described Table 2 5 from the work of Grossman (1970) sets this out

Further, as the effects of these four hormones were explored, their targets became more and more numerous Now the list includes *1)* water and electrolyte secretion by stomach, pancreas, liver, and Brunner's glands, *2)* enzyme secretion by stomach and pancreas, *3)* water and electrolyte absorption by ileum and gall bladder, *4)* release of insulin, glucagon, secretin, and gastrin, *5)* smooth muscle of lower esophageal sphincter, stomach, small and large

Table 2 3

Structures of gastrointestinal hormones[a] The structure of secretin and gastrin have been confirmed by total synthesis Synthesis of the active C-terminal heptapeptide amide of cholecystokinin has also been accomplished

HUMAN GASTRIN II (17 residues)

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Pyr-	Gly-	Pro-	Trp-	Leu-	Glu-	Glu-	Glu-	Glu-	Glu-	Ala-	Tyr-	Gly-	Trp-	Met-	Asp-
											HSO_3				

17	
Phe-	NH_2

PORCINE CHOLECYSTOKININ (33 residues)

Lys-	(Ala,	Gly,	Pro,	Ser)-	Arg-	Val-	(Ile,	Met,	Ser)-	Lys-	Asn-	(Asn,	Gln,	His,	Leu_2,
Pro,	Ser_2)-	Arg-	Ile-	(Asp,	Ser)-	Arg-	Asp-	Tyr-	Met-	Gly-	Trp-	Met-	Asp-	Phe-	NH_2
								HSO_3							

PORCINE SECRETIN (27 residues)

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
His-	Ser-	Asp-	Gly-	Thr-	Phe-	Thr-	Ser-	Glu-	Leu-	Ser-	Arg-	Leu-	Arg-	Asp-	Ser-

17	18	19	20	21	22	23	24	25	26	27	
Ala-	Arg-	Leu-	Gln-	Arg-	Leu-	Leu-	Gln-	Gly-	Leu-	Val-	NH_2

[a] Parentheses indicate positions of cholecystokinin with unresolved amino acid sequences

Table 2 4
Carboxy-terminal heptapeptide amides The residues are numbered unconventionally that is starting from C terminus

	7	6	5	4	3	2	1	
Gastrin II	Ala-	Tyr-	Gly	Trp-	Met	Asp-	Phe-	NH_2
		HSO_3						
Cholecystokinin	Tyr-	Met-	Gly	Trp-	Met-	Asp-	Phe-	NH_2
	HSO_3							

Table 2 5
Each of the three pure gastrointestinal hormones has an effect on each of the four primary actions by which the hormones were first identified[a]

Hormone	Gastric H	Pancreatic HCO_3	Pancreatic Enzymes	Gallbladder
Secretin	↓↓	↑↑	↑	↑*
Gastrin	↑↑	↑	↑↑	↑
Cholecystokinin	↑-↓	↑	↑↑	↑↑

[a] ↑, stimulates, ↑↑, stimulates strongly, ↓, inhibits, ↓↓, inhibits strongly, ↑-↓, stimulates alone and inhibits when given with gastrin, ↑*, stimulates when given with cholecystokinin

bowel, gallbladder, and sphincter of Oddi and uterus, *6*) release of histamine and augmentation of histadine decarboxylase activity in gastric mucosa, *7*) amino acid uptake by gastric mucosa and pancreas, *8*) growth of gastric mucosa and pancreas, *9*) some metabolic actions lipolysis and glycogenolysis, and *10*) cardiovascular actions including inotropic and chronotropic actions on the heart and effects on blood flow in bowel, liver, gastric mucosa, pancreas, and extremities Table 2 6, from the pioneer work of M I Grossman (1970), details the stimulating and inhibiting actions of these four hormones as they affect secretion, absorption, and muscle contraction

To order this complex area of data, Grossman has hypothesized that the target organs contain a receptor complex with two units —one reactive to gastrin and cholecystokinin, and the other to glucagon and secretin—and that the two sites interact

The great unsolved problem is whether the effects on sites of other than the primary or identifying organ have any "physiological" significance, since in some instances, large doses must be employed In many, however, the dose range seems to fall in the "physiological" one

In addition to the interaction of hormone

Table 2 6
A catalog of actions of four hormones[a]

	Gastrin	Cholecystokinin	Secretin	Glucagon
Water electrolyte secretion				
Stomach	↑	↑	↓	↓
Pancreas	↑	↑	↑	↓
Liver	↑	↑	↑	↑
Brunner's glands	↑	↑	↑	↑
Water electrolyte absorption				
Ileum	↓	↓	↓	NT
Gallbladder	0	0	↓	NT
Enzyme secretion				
Stomach	↑	↑	↑	0
Pancreas	↑	↑	↑	↓
Endocrine secretion				
Secretin	↓	NT	NT	NT
Gastrin	NT	NT	↓	NT
Insulin	↑	↑	↑	↓
Glucagon	0	↑	0	NT
Smooth muscle				
Lower esophageal sphincter	↑	NT	↓	NT
Stomach	↑	↓	↓	↓
Intestine	↑	↑	↓	↓
Ileocecal sphincter	↓	NT	NT	NT
Gallbladder	↑	↑	↑	NT
Sphincter of Oddi	↓	↓	NT	↓
Uterus	↑	NT	NT	NT

[a] ↑, stimulates, ↓, inhibits, 0, no effect, NT, not tested

with hormone so far discussed, interactions between neural effects and hormones on the gut are likewise important, especially *vagal* effects These are of two main kinds, the vagal release of hormones (especially in stimulation of acid secretion) and vagal effects on the action of the hormones on the target cells (especially the gastric glands), but there is no reason not to think that these interactions hold for the remainder of the gut

4 / SALIVARY SECRETION

The Salivary Glands

Only mammals possess true salivary (digestive) glands Oral mucous glands, however, are found in all terrestrial vertebrates There is considerable variation in the number of salivary glands and also in the cellular structure of individual glands In addition, the products of secretion by salivary glands vary not only from species to species, but from one salivary gland to another in the same animal

In man, salivary secretion is provided by three pairs of glands, the *parotid*, *submaxillary*, and *sublingual* glands Also located within the oral and pharyngeal cavities are numerous other smaller glands that contribute to the production of saliva These smaller glands include the lingual mucous glands located along the borders and at the root of the tongue, lingual serous (Ebner's) glands located in the vicinity of the vallate and foliate papillae of the tongue, and numerous small mucous glands in the mucous membrane covering the cheeks, hard and soft palate, and the walls of the pharynx

The production of saliva by the salivary glands provides the first digestive juice that comes in contact with food In addition to containing salivary amylase that initiates starch digestion, the salivary secretions play an important role in oral hygiene, the sensitivity of taste perception, the process of deglutition, and also serve as a medium for the excretion of certain inorganic and organic substances

GENERAL FORM AND LOCATION

The *parotid glands* are the largest of the salivary glands and in the adult weigh from 20 to 30 g each The parotid gland is located at the side of the face below and in front of the ear

This gland is enclosed in a sheath which limits the swelling of the inflamed gland Because of this facial presentation of the gland, relatively small degrees of swelling are often conspicuous, as in mumps As the anteromedial surface of the gland is in intimate contact with the masseter muscle and the ramus of the mandible, movement of these structures when chewing usually aggravates the pain and tenderness of a swollen parotid gland

The *duct system* of the parotid gland originates in the *acini* of the gland Ducts from neighboring acini join to form ducts of larger caliber which unite again to form still larger ducts The fine ducts that drain the acini are known as *intercalated* ducts The intercalated ducts empty into *intralobular* ducts and these into *interlobular* or excretory ducts Thus, by a progressive convergence of the several orders of ducts, the gland's secretion is conducted finally into a single large duct which opens into the oral cavity This general arrangement of the ductal system is similar to the stem branching of a bunch of grapes—the rounded acini at the ends of the finest channels corresponding to the grapes The term *racemose* (Latin for bunch of grapes) has been applied to glands exhibiting such a structure, and *acinus* (Latin for grape) identifies the smallest group of secretory cells of such a gland

The parotid duct (*Stensen's duct*) emerges from the anterior border of the gland near its upper end and runs transversely across the surface of the masseter muscle At the anterior border of the masseter muscle, the parotid duct abruptly turns medialward, pierces the buccal fat pad and the mucosa to open into the oral cavity at a point lateral to the crowns of the second upper molar The parotid duct is approximately 35 to 40 mm long and about 3 mm in diameter The wall of the duct is tough due to the presence of a thick layer of fibrous tissue

The *submaxillary glands* are shaped somewhat like a flattened walnut and weigh 8 to 10 g each They are located in the submaxillary triangle medial to the mandible Like the parotid glands, the submaxillary glands have a racemose structure The duct system is similar to that of the parotid glands and collects in a large, single submaxillary duct known as *Wharton's duct* The submaxillary duct emerges from the deep surface of the gland and runs forward

and inward to reach the floor of the mouth It opens by a small orifice on the summit of a soft papilla known as the *caruncula sublingualis* by the side of the frenulum of the tongue The submaxillary duct is 40 to 50 mm in length and the walls of the duct are relatively thin

The *sublingual glands* are the smallest of the paired salivary glands and weigh from 2 to 3 g each The sublingual glands are elongated in shape and lie immediately under the mucous membrane of the floor of the mouth under the tongue The structure of the sublingual gland is also racemose Instead of being drained by a single, large duct, the sublingual glands are drained by 5 to 15 or more small, short ducts (*ducts of Rivinus*) that arise from the upper border of the gland and open on small papillae along the crest of the *plicae sublingualis* The anterior portion of the gland is often drained by a single, larger duct, also known as *Bartholin's duct*, which opens alongside the submaxillary duct on the sublingual caruncle Unlike the parotid and submaxillary glands, the capsule of the sublingual glands is poorly developed

MICROSCOPIC ANATOMY

The racemose structure of the salivary glands is found on histological section to consist of acini, a tubule system, and excretory ducts The secreting epithelium of the acini is composed of two types of cells, *mucous* and *serous*, that can be differentiated from one another on the basis of nuclear appearance and staining reactions The nucleus of the mucous cell is small, dark-stained, indented and compressed against the lower pole of the cell The serous cell possesses a larger, centrally placed vesicular nucleus Four different types of cells may be distinguished in the lining of the salivary ducts In addition, *myo-epithelial* cells which have been described in detail by Zimmerman (1927) have been demonstrated between the basal membrane of the acini and ducts

Acinar Structure The acini of the salivary glands are composed of cells that are more or less pyramidal in shape The apices of adjacent cells of an acinus converge upon a minute central lumen, the basal aspects of the cells form the periphery of the acinus The central lumen is continuous with that of the excretory duct system

The acini of the parotid gland in *man* consists entirely of serous cells. Parotid secretion collected separately from the other salivary glands has a characteristic watery character or serous secretion In man, submaxillary gland serous and mucous elements are seen in a ratio of about 4 1 More often than not the mucous acini are "capped" by a single layer of flattened serous cells, referred to as serous demilunes (Fig 2 8) The secretion of the serous demilune is believed to reach the lumen of the acinus by means of small canaliculi between the mucous cells The acinar structure of the sublingual glands is similar to the submaxillary gland except that the ratio of serous to mucous elements is about 1 4

The mucous cells have large translucent granules which readily transmit light, so that in thin sections they appear transparent The serous cells have small granules that are more opaque and appear darker in sections than the mucous cells The serous cells secrete a thin, watery fluid containing the starch-splitting enzyme, amylase, also known as *ptyalin*

The myo-epithelial, or basket cells, are flat

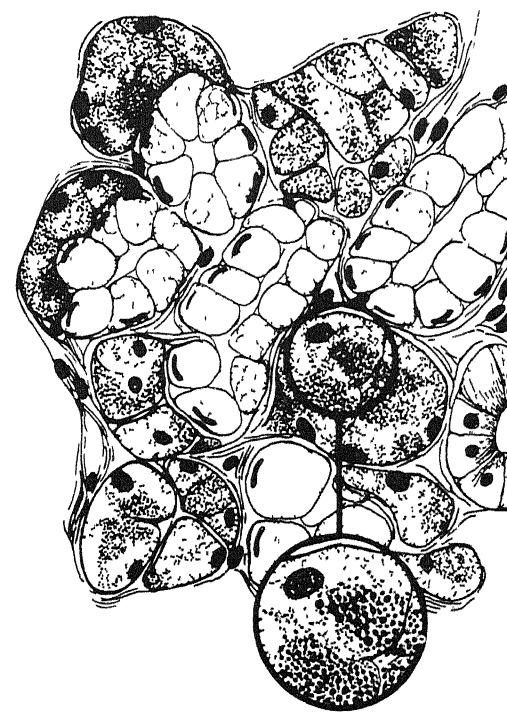


Fig 2 8 Microscopic section of human submaxillary gland showing both mucous and serous groups of cells The mucous cells are lightly stained and appear transparent The serous cells are darker and some appear as demilune cells in the upper left area of the microscopic section The insert is an oil immersion view showing zymogen granules in serous cells (From Stormont 1932)

and somewhat star-shaped with long processes, and, as previously mentioned, lie between the basement membrane and the secretory cells of the acinus They are also found in the intralobular salivary ducts between the basement membrane and the cells lining the lumen The myo-epithelial cells are modified epithelial cells and are thought to have a contractile function, although there is no clear evidence of this The myo-epithelial cells of the salivary glands may serve a function similar to the myo-epithelial cells of the mammary gland In the mammary gland, these cells are well developed and it has been established that their contraction results in the expulsion of preformed milk from the ducts This is observed in the lactating breast and is known as "let down " It has been observed that stimulation of the sympathetic nerve supply to the salivary glands sometimes causes a sudden gush of saliva from the gland This suggests that some previously secreted saliva is being expressed from the ducts by contraction of the myo-epithelial cells

The cytoplasm of both the serous and mucous cells show a glandular structure The fine granules of the serous cells are believed to furnish the enzymes of the secretion and are called zymogen granules The mucinogen granules of the mucous cells provide the mucin that renders the secretion viscous When the glands are in a resting state, the granules accumulate and the fine zymogen granules may load the cytoplasm to such an extent that the nucleus is almost obscured The concentration of secretory granules, mucous as well as serous, is low soon after meals and gradually increases prior to the next meal The *ultrastructure of the serous cells* of the salivary glands has been found to be quite similar to the acinar cells of the pancreas

Nerve fibers have been observed penetrating the basement membrane of a salivary acinus and ending as fine filaments on the basal surface of the acinar cells (Scott and Pease, 1957) The electron microscope reveals the myo-epithelial cell to have a close resemblance to a smooth muscle cell (Porter, 1954) This similarity tends to support the theory that the myo-epithelial cells have a contractile function

Ductal Structure Four different types of cells are found in the salivary ducts In the *intercalary ducts* that lead directly away from the alveoli, the epithelium is *cuboidal* in character These small cuboidal cells are practically filled by a large nucleus The cytoplasm is scanty and rarely are any cytoplasmic granules observed The *intralobular ducts* are lined by *columnar* cells that present a peculiar rod-shaped appearance, resembling in this respect, some of the cells of the renal tubules These cells have a centrally placed nucleus and there is a marked striation of the basal one-third of the cell Electron micrographs reveal that the marked striation observed by the light microscope in the basal third of the columnar cells lining the intralobular ducts is due to the alignment of mitochondria by infoldings of the plasma membrane This arrangement, which greatly increases the surface area of the cell, suggests that it facilitates the transport of materials into or out of the cells The *interlobular* or excretory ducts of the salivary glands are lined by a *two-layered epithelium* composed of a *columnar* superficial layer and a *flattened* deep layer This type of lining of the excretory ducts is continued to near the termination of the duct, where it empties into the mouth At this point, the lining of the duct changes to a layered, *stratified, squamous epithelium* The fact that the cells that line the various portions of the salivary ducts differ in structure suggests that their functions may be different The rod-shaped epithelium of the intralobular ducts is particularly suggestive of a secretory or reabsorptive function

INNERVATION OF THE SALIVARY GLANDS

The salivary glands are supplied with efferent nerves from both the *parasympathetic* and *sympathetic* divisions of the *autonomic nervous system* Functions regulated by the parasympathetic nerve supply are secretion by the acinar cells and vasodilation of the blood vessels Functions regulated by the sympathetic nerve supply include vasoconstriction of the vascular supply, and, presumably, contraction or motor function by the myo-epithelial cells

The origin of the *parasympathetic nerve supply* is in *bulbar centers*, which consist of a group of nerve cells which run forward from the anterior end of the glossopharyngeal nucleus to the sensory nucleus of the facial nerve The anterior (rostral) part is termed the *superior salivary nucleus*, and governs secretion by the submaxillary and sublingual glands, the posterior (caudal) part, called the *inferior salivary nucleus* controls secretion by the parotid gland

The *submaxillary and sublingual glands* receive secretory impulses through the *chorda*

tympani nerve The preganglionic fibers carried by the chorda tympani end in the submandibular ganglion by making synapse with the cell bodies of the neurons of the ganglion The axons arising from the latter constitute the postganglionic fibers which terminate by fine arborizations around the serous cells of the submandibular gland (Figs 2 9 and 2 10)

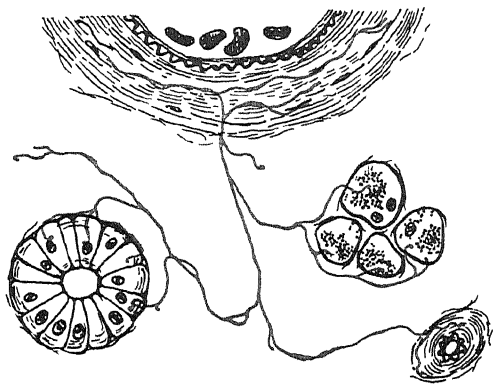


Fig 2 9 Distribution of parasympathetic secretory nerves to cells and blood vessels of submaxillary gland of rabbit (From Stormont 1932)

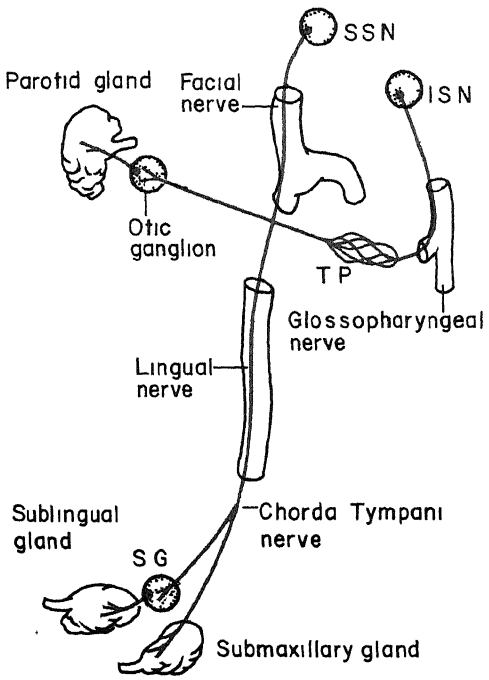


Fig 2 10 Diagram of the parasympathetic nerve supply of the salivary glands S S-N superior salivatory nucleus I-S-N inferior salivatory nucleus S-G submaxillary ganglion T-P tympanic plexus

The parasympathetic or bulbar fibers to the *parotid* gland arise in the medulla from the *inferior salivary nucleus* They pursue the first part of their course in the *glossopharyngeal nerve* The fibers then are conveyed to the *otic ganglion* There they communicate with ganglion cells from which postganglionic fibers arise that innervate the acinar cells

The *sympathetic nerve supply* (preganglionic fibers) to the salivary glands is derived from the lateral horn within the first and second thoracic segments of the spinal cord Axons from the lateral horn cells leave the cord via the ventral roots, enter the paravertebral sympathetic trunk, and continue craniad to the *superior cervical ganglion* where synapse is made with the cell bodies of the second order neuron From the latter, axons arise and follow the external carotid artery to the parotid gland Whether or not these fibers innervate more than vascular musculature is a moot question

Afferent nerve fibers from the salivary glands are found in the *chorda tympani* and the *glossopharyngeal nerves* These fibers carry pain impulses from the salivary glands Claude Bernard (1858) showed that stimulation of the submaxillary gland in a nonanesthetized dog caused evidence of pain Swelling of the gland, as occurs in mumps or when the main salivary duct is obstructed, is also known to be painful

BLOOD SUPPLY AND LYMPHATIC DRAINAGE

Blood flow within the salivary glands, as we shall see later, has an intimate relationship to the process of salivary secretion The arterial supply to the parotid gland is from the *posterior auricular* artery The arteries to the submaxillary gland are derived from the *external maxillary* (facial) and *lingual* arteries, while the arterial supply to the sublingual gland is supplied from the *sublingual* and *submental* arteries In the case of the submaxillary and sublingual glands, the main arterial supply enters the gland at the hilus The parotid gland does not have a well-defined "hilus," and, correspondingly, point of entrance of its arterial supply is subject to variation

Within the gland substance, the arterial supply travels with the interlobular ducts into the lobules Within the lobules, the arteries break up to form a rich capillary plexus about the ducts Consistent with the observation that the ducts perform active transport functions is the fact that the vascularization of the ducts is

much richer than that of the acini Arteriovenous anastomoses similar to the juxtaglomerular vessels in the kidney have been described by Spanner (1937) The function of such arteriovenous anastomoses is not known The capillary plexuses about the alveoli and ducts collect in venules which, in general, accompany the ducts out of the lobules Near the hilus of the gland, Spanner has described large sacculated veins which he suggests serve a reservoir function and tend to back up pressure in the capillary circulation during secretory activity Because of this suggested function, these sacculated, venous structures have been referred to as "throttle" veins Similar veins are found in the penis In studies on the circulation within the salivary glands, Burgen and Seeman (1958) have found that most of the blood perfusing the gland went to capillaries in the duct system and that the flow in these was mainly countercurrent to the flow of the saliva

Lymph spaces are found around the acini which drain into lymphatics that accompany the ducts and vessels Lymphatic drainage from all of the salivary glands is into the superficial and deep cervical glands

Formation of Saliva

The mechanisms of salivary secretion that result in the formation of saliva have been studied extensively but still are only incompletely understood It was thought that the formation of saliva was accomplished by processes similar to those that result in the production of urine by the kidney It was assumed that saliva was an ultrafiltrate of the blood The ultrafiltration theory was disproved by Ludwig (1851) when he showed that the submaxillary gland could secrete against a pressure greater than its arterial blood pressure

It is now recognized that the mechanism of salivary secretion is one that requires energy for the production and secretion of organic substances by the acinar elements and for the "active transport" of inorganic substances across cell membranes against concentration gradients by the duct epithelium

The evidence that salivary secretion is a process that requires energy has been accumulated in a number of ways Northup (1935) has shown that the salivary gland utilized a metabolic process to obtain energy for the work of secretion that is in many ways similar to that of muscle

The *respiratory quotient* (RQ) of slices of maxillary gland suspended in Ringer's solution was found by Deutsch and Raper (1938) to be between 0 6 and 0 8 When they added acetylcholine and eserine to the Ringer's solution, the rate of respiration was increased, but in the absence of glucose, the RQ did not change If glucose was added to the slices during incubation, the RQ increased to 1 0 Epinephrine had an effect similar to that of acetylcholine

Barcroft (1914) found oxygen consumption for the resting submaxillary gland of the dog to be from 22 to 27 μl /g /min In the submaxillary gland of the cat, Stromblad (1957) obtained a value of 20 μl /g /min When the gland is stimulated by injecting acetylcholine or by parasympathetic stimulation, the oxygen consumption increases to 130 to 175 μl /g /min Atropine can block the metabolic effects of acetylcholine Metabolic poisons such as cyanide, iodoacetate and fluoride are also capable of blocking the metabolic effect of acetylcholine

It has also been shown (Hokin and Hokin, 1959) that acetylcholine and epinephrine will increase the incorporation of radioactive phosphorus into the phospholipids of the rabbit parotid and submaxillary and in guinea pig parotid gland slices Atropine and ergotamine are capable of blocking this effect A large proportion of the total respiration of the salivary glands must be attributed to the duct epithelium since these cells possess mitochondria in very high concentration

It is now recognized that the intralobular ducts contribute considerably to the formation of saliva Since saliva is hypotonic to blood serum, this must be accounted for either by the reabsorption of electrolytes in the ducts or by the lack of transfer of electrolytes into the saliva across the cellular membrane To elucidate the role of the salivary ducts in the formation of saliva, Burgen and Emmelin (1961) devised an ingenious method to demonstrate the site of entry of various substances into the saliva by the ductal epithelium The method consisted of rapidly injecting a small amount of an isotope into the arterial blood supply of the salivary gland when the gland is secreting at a steady state The appearance of the isotopic substance in the collected saliva is then determined The amount of isotope used is quite small and thus the secreting gland is exposed to the isotope only briefly After passage through the gland, the isotope is diluted in the animal's blood and body fluids

The appearance time of the isotope in the

saliva is only a matter of seconds and is an indication of the site of its passage across the ductal epithelium to enter the saliva Thus, the more rapid the appearance of the substance in the saliva, the larger the order of duct representing the site of transfer When two or more isotopic substances are injected simultaneously, the same relationship holds, i e , the substance that appears first in the saliva was transported by the ductal epithelium at a point along the duct system that is distal to that at which the other substance was transported Using this technique, the appearance time of water, chloride, bromide, iodide, sulfate, bicarbonate, sodium, potassium, urea, and four amino acids have been determined in parotid saliva of the dog None of the four amino acids, valine, methionine, isoleucine and tyrosine, nor sulfate were detected in any appreciable amounts in the saliva All of the other isotopes appeared after a more of less brief latent period of 2 to 10 sec Following the appearance of the isotopic substance, its concentration rapidly increased to a peak in 5 to 6 sec , then gradually declined during the next 2 to 3 min (Fig 2 11)

Isotopic water appeared first in the saliva, indicating that it was capable of being transported across the ductal epithelium at a more distal site than any of the other substances

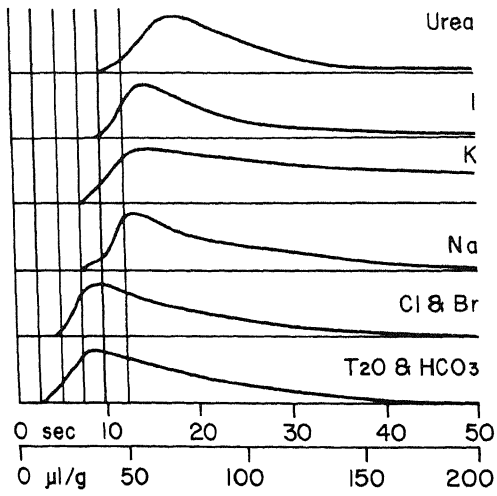


Fig 2 11 Secretion patterns of various isotopic substances after injection into the arterial blood supply of the parotid gland of the dog The appearance of the isotopes was determined in single drops of saliva secreted by the gland Note that tritiated water and bicarbonate appeared first and urea last Except for potassium the outflow curves of all other substances were similar (re-drawn from Burgen and Emmelin 1961)

Bicarbonate was found to enter the saliva at the same point as water and had a similar time concentration pattern Chloride and bromide appeared next, indicating that they entered the saliva through the epithelium of somewhat smaller ducts than those of water and bicarbonate Sodium and potassium ranked third in order of appearance, iodide and urea, fourth It will be noted from Figure 2 11 that the declining slope of the potassium curve after it reaches a peak is quite slow This slow decline of the potassium concentration is taken as evidence that potassium is also secreted into the saliva by the acinar cells

The fact that anions can enter the ducts at a more distal site than cations suggests that anionic exchange may occur at this point If cations are unable to enter the duct at some point, the only explanation of anion entry would be by an anionic exchange mechanism It will be discussed later that there is a reciprocal relationship between bicarbonate and chloride that is controlled by the arterial P_{CO_2} The bicarbonate-chloride exchange may occur in the more distal part of the salivary ducts

Composition of Saliva

The composition of saliva consists mainly of water, inorganic constituents commonly found in the plasma, and certain organic constituents The variability of the composition of saliva is accounted for by the fact that the different salivary glands contribute different constituents and the final composition is adapted to stimuli evoking the secretion In addition, the concentration of many of the common inorganic constituents depends upon flow rate

VOLUME AND FLOW RATE

In man, the amount of saliva secreted in 24 hr amounts to 1000 to 1500 ml (about 1 ml /min) A large proportion of this 24-hr volume is secreted at mealtime, when the secretory rate is highest In the absence of stimulation, as during sleep, the secretory rate is low and may not exceed 0 25 ml /min Schneyer and Levin (1955) have determined that under "resting" conditions the parotids contribute approximately 25%, the submaxillary gland 70%, and the sublinguals 5% In the dog, the maximum secretory rate of the parotid gland is approximately 0 5 ml /g of gland/min , whereas the maximum rate of the submaxillary gland is approximately 1 ml /g of gland/min

Ordinary mixed saliva contains about 99 5% water and 0 5% total solids The specific gravity of mixed saliva is usually between 1 002 and 1 012 The osmolarity of saliva from all of the salivary glands in man is slightly hypotonic to that of plasma Thus, some osmotic work is required for the secretory process The osmolarity of saliva can be increased by dehydration and decreased by subsequent hydration

REACTION

Human mixed saliva is slightly acid in reaction In a large series of normal persons, Starr (1922) found the pH of mixed saliva to vary between 5 75 and 7 05 The pH of saliva is dependent mainly upon the relative concentrations of free and combined CO_2, that is, upon the ratio of $H_2CO_3/NaHCO_3$ Thus, for the true pH value to be obtained, the juice must be collected without loss of CO_2 The older figures in the literature of 7 50 to 8 00 for the pH of saliva are too high, the precaution to prevent loss of CO_2 was not taken The pH of the saliva varies directly with the CO_2 content of the blood This means that when the CO_2 tension in the blood is high, more CO_2 finds its way into the salivary secretion to lower its pH and vice versa Force breathing causes a lessened amount of CO_2 in the saliva and, consequently, a rise in its pH

The *bicarbonates*, and to some extent, the *phosphates*, act as "buffers" in saliva This buffering action is such that the pH of saliva remains relatively constant under all ordinary conditions Even when relatively strong solutions of acid or alkali are introduced into the mount, the reaction, except for such a short period, is not altered

INORGANIC CONSTITUENTS

Bicarbonate concentrate in saliva increases with salivary flow At low rates of secretion, the bicarbonate concentration may be from 5 to 10 mEq /liter With increased rates of secretion, the bicarbonate concentration *exceeds* that of the serum In human parotid saliva, the bicarbonate concentration may go as high as 60 mEq /liter Sublingual saliva has a low bicarbonate concentration Salivary bicarbonate is derived from both gland metabolism and the transfer of bicarbonate from the plasma to the saliva

With changes in bicarbonate concentration, there is a reciprocal change in the chloride concentration and the total anionic concentration is not markedly altered Thus, with an increased arterial P_{CO_2}, bicarbonate enters the saliva and exchanges for chloride The site of this anionic exchange is probably in the more distal parts of the striated salivary ducts

Phosphate concentration in saliva is about twice that of plasma and relatively independent of flow rate Approximately 80% of the phosphate is present as inorganic phosphate The organic phosphate in saliva is acid-soluble and partially split by phosphates

Chloride concentration of saliva from different glands may show considerable variation In general, there is nearly a linear relationship between the chloride concentration and flow rate (see Fig 2 12) At all flow rates, however, the salivary chloride concentration is *below* that of plasma and may vary between 5 and 70 mEq /liter An exception to this general observation is sublingual saliva from the cat, which has a higher chloride concentration than that of plasma Sublingual saliva of man has a higher chloride concentration than parotid or submaxillary saliva

In patients with salt depletion, there is usually little change in the salivary chloride Thus, the salivary chloride concentrations tends to be independent of plasma chloride concentration

Salivary secretion of *bromide* is similar to that of chloride However, the salivary/plasma ratio tends to be slightly higher Little is known about the secretion of *fluoride* by the salivary glands except that it is secreted in the saliva at a lower concentration than the other halides McClure (1941) has reported that the salivary

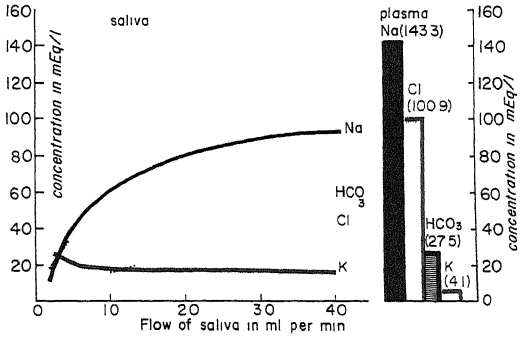


Fig 2 12 Relationship between concentrations of sodium potassium chloride, and bicarbonate in human parotid saliva and rate of salivary flow Plasma concentrations of the ions are also shown (From Thaysen et al 1954)

concentration of fluoride in children is from 0 006 to 0 015 mEq /liter Because of the relationship of fluorides to dental caries, more information is needed on the salivary secretion of fluoride

Sodium concentration in saliva is highly dependent upon the rate of saliva flow At low flow rates, the concentration may be less than 5 mEq /liter, whereas at high flow rates, it may exceed 100 mEq /liter Thus, it is impossible to give a "normal" sodium concentration for saliva Resting saliva in normal children contains from 2 to 15 mEq /liter of sodium In children with cystic fibrosis of the pancreas, the sodium concentration is increased and reaches levels from 20 to 45 mEq /liter

Potassium concentration of saliva is relatively high and exceeds that of the blood plasma All of the main salivary glands of man secrete potassium at a concentration higher than that of plasma Ordinary mixed saliva usually contains 8 to 20 mEq /liter of potassium which is 1 5 to 4 times the concentration in plasma Values as high as 50 to 60 mEq /liter have been found in the submaxillary and sublingual saliva of rats The potassium concentration of saliva resulting from sympathetic stimulation is about twice that obtained with parasympathetic stimulation

The concentration of *calcium* in human parotid saliva, like sodium, increases with flow rate At high flow rates, a concentration of 3 to 4 mEq /liter may be found

ORGANIC CONSTITUENTS

Salivary proteins account for the physical characteristics of saliva Parotid saliva has a low viscosity and is watery in character, whereas submaxillary saliva usually is viscous and sticky due to its mucoprotein content With increasing flow rates, the concentration of protein increases However, with prolonged stimulation, the quantity of protein secreted per unit of time progressively decreases This suggests that protein production by the salivary glands is relatively slow compared with the rate of protein secretion Stimulation of the sympathetic nerves to the submaxillary gland of the cat or dog produces a saliva rich in proteins The concentration and character of the proteins differ from that obtained by parasympathetic stimulation

The proteins are extremely complex in nature and the older classification into albumins and mucins is inadequate Electrophoretic, ultracentrifugation and immunological methods have been used in attempts at identification and classification of salivary proteins The electrophoretic pattern of salivary proteins differs greatly from that of serum In human parotid saliva, the main electrophoretic peak is *amylase* In submaxillary saliva, the major component is a *mucoprotein* of low mobility

Free amino acids are found in saliva Their concentrations are usually 10 to 20% of the plasma levels When plasma levels of free amino acids are raised, the concentration of the amino acids increase in saliva *Urea, uric acid* and *creatinine* are also found in saliva

Amylase is the main digestive enzyme of parotid saliva in man Amylase is also found in the saliva of apes, rabbit, guinea pig, rat and mouse Very little amylase is found in parotid and submaxillary saliva of the dog, cat, cow, horse, sheep, and goat Apparently, starch digestion by salivary amylase is not too important in some animals Amylase can be crystallized from saliva by successive acetone and ammonium sulfate precipitation The optimum pH of amylase is 6 9 to 7 1 and requires chloride ions for its activation The Ca^{++} ion is also required for the activation of amylase Each mole of amylase contains 1 gram atom of calcium, which, if removed, inactivates the amylase

On acrylamide electrophoresis, crystalline salivary amylase has been demonstrated to consist of a number of isoamylases There is usually a major amylolytic band and up to five minor amylolytic bands The number of isoamylases in parotid saliva may vary from three to six in different individuals (Hall, Gulig, Hightower, 1969, Boettcher and de la Lande, 1969) The isoamylases have a molecular weight of approximately 50,000

Lysozyme is an interesting enzymatic substance found in saliva This enzyme, which appears to be a mucoprotein, has the ability of lysing certain bacteria Lysozyme is bacteriolytic for organisms belonging to the genera *Bacillus, Micrococcus, Staphylococcus, Streptococcus, Proteus,* and *Brucella*

Two additional organic constituents of saliva that are of interest are *kallikrein* and the specific *blood group substances* When saliva is injected intravenously into an animal, it produces a fall in blood pressure for a brief period The substance responsible for this has been

found to be due to kallikrein, an enzyme that acts on plasma protein to produce a vasodilator polypeptide This polypeptide is known as *kallidin* or *bradykinin* It has been suggested that the vasodilation in the salivary gland during secretory activity is due to kallikrein release The soluble specific blood group substances found in saliva have the same characteristics as the agglutinogen on the erythrocyte These substances can be demonstrated by the direct precipitin reaction with type specific sera In man, the A, B, O and Le^a substances have been demonstrated The concentration of the blood group substances in saliva is from 10 to 20 mg /liter It appears that most of the blood group substances come from the mucous cells of the submaxillary gland whereas little is produced by the parotid

Glucose is normally absent from saliva and even in patients with diabetes, none or only small amounts are found

Control of Salivary Secretions

Saliva is secreted continuously, but the rate of secretion may vary greatly due to the influence of conditions that are capable of stimulating or inhibiting the process of secretion The *continuous secretion* of saliva in the absence of any known stimulating factors is called "spontaneous" or "resting" secretion The continuous or spontaneous secretion of saliva is important in keeping the mucous membrane of the mouth and the pharynx moist at all times Examples of spontaneous or continuous salivary secretion in animals include the parotid gland of the sheep and ox and the sublingual gland of the cat The mechanism of spontaneous secretion is not known but it has been suggested that it is due to the constant release of minute amounts of acetylcholine within the gland However, atropine has no effect on this type of secretion

During mastication of food the rate of salivary secretion is greatly increased The character of the secretion is adapted in some degree to the type of food undergoing mastication During sleep the rate of salivary secretion is low and accounts for the dry mouth many experience on waking up in the morning Sexual excitement in many animals is associated with an increase in salivary flow

There are two main ways of controlling secretion of the digestive glands one through the nervous system and the other by means of hormones In the case of the salivary glands, it is believed that the control is exclusively through the nervous system At any rate, all efforts to date to find a salivary hormone have failed

REFLEX SECRETION OF SALIVA

The secretion of saliva is brought about reflexly in two ways either through the stimulation of the nerves of the mouth by the presence therein of food or other substances, or by the stimulation of some organ of special sense other than that of taste The former type of reflex is termed *unconditioned* or *inherent*, the latter, *conditioned* or *acquired* A reflex of one type does not, of course, exclude the other and, as a matter of fact, both are called into play together under ordinary circumstances

In addition to the unconditioned salivary reflex brought about by materials placed in the mouth, the salivary glands can be excited readily by the artificial stimulation of afferent nerve supplying structures in the mouth The conditioned salivary reflex depends upon afferent impulses arising from outside the oral cavity

The Unconditioned Salivary Reflex. Materials placed in the mouth call forth, after a short latent period (2 to 3 sec), a secretion of saliva which varies in *quantity* and *quality* with the physical and chemical nature of the substance introduced The effect which sensations of taste produce upon the secretion of saliva are well known Among edible substances, those, generally speaking, which are the most palatable or arouse the sensation of taste with the greatest intensity are the most potent salivary stimulants Materials, especially acids, which are entirely inedible will, if unpleasant to the taste, cause profuse salivation In these instances, the secretions depend mainly upon the stimulation of the taste fibers, and the stimulus is chemical in nature Materials such as dry sand, inedible powders (whether soluble or insoluble) or any other material which is capable of stimulating in a purely physical way will evoke a secretion. The mere movements of the jaws and of the tongue over the mucosa of the mouth will have such an effect A secretion occurs when any substance is chewed, whether or not it is edible or possesses taste The manipulations of the dentist, the contact of his instruments with the oral mucosa, or the grinding of a tooth are familiar and effective salivatory stimuli The salivary glands thus can

be readily excited by the artificial stimulation of afferent nerves supplying structures in the mouth

The reflexes are brought about through the salivary centers The *efferent* limbs of the reflex arcs, including the secretory fibers of the chorda tympani and the tympanic branch of the glossopharyngeal nerves The *afferent* pathways are represented in the trunks of the chorda tympani, the pharyngeal branches of the vagus and glossopharyngeal nerves, and the lingual, buccal, and palatine branches of the trigeminal nerve

The Conditioned or Acquired Reflex. The secretion which flows into the empty mouth when "the mouth waters" is a conditioned reflex The stimulus which initiates such a reflex is not applied to the nerves of the mouth, but is received by one or other of the organs of special sense, particularly those of sight and smell A conditioned reflex may also be elicited through the sense of hearing, or through sensory impressions arising from stimuli applied to the skin In brief, a conditioned reflex is one in which the cerebral centers play an essential part, and in which training and experience are the basis for the development of the reflex process

ADAPTABILITY OF THE SALIVARY REFLEX

It has been mentioned previously that saliva varies in both quantity and quality with the physical and chemical nature of the substance stimulating the secretion This remarkable adaptability or purposeful character of the salivary reflex was observed by Pavlov (1910) and later confirmed by Babkin (1950) The physical and chemical qualities of the juice, to some extent, are conditioned by the agent evoking salivation For instance, if clean pebbles are placed in a dog's mouth, they are expelled—merely allowed to drop out of their own weight Little or no secretion occurs, since none is required, but if the stones are crushed and given as a powder, a profuse water salivation follows to rid and cleanse the mouth of the useless material The juice, in this instance, is poor in organic material and resembles that obtained upon electrical stimulation of the chorda tympani nerve or of the parasympathetic fibers to the parotid Strong acid produces an abundant saliva which, according to Pavlov, is relatively rich in proteins, this exerts a buffer action which reduces or annuls the injurious effect of the acid The salivary response provides a lubricant which is curiously appropriate in volume and viscosity for the food to be swallowed A chunk of meat, given to a dog, is very quickly coated with a viscous mucin-rich saliva and swallowed If the meat is first dried and powdered, or if a dried biscuit is fed to the animal, an abundant watery saliva is produced such as is characteristic of parotid saliva

PARALYTIC SALIVARY SECRETION

Claude Bernard observed in 1864 that if he cut the chorda tympani nerve in a dog or a cat, in about 24 hr the submaxillary gland began to produce a scanty secretion of thin, turbid saliva, the turbidity was due to the presence of leukocytes The secretion increased until the 7th or 8th day, at which time it reached a plateau and began to diminish about the 3rd week It stopped some 6 weeks after the nerve was cut During this time the gland increased in size, and no secretion could be obtained on stimulation of the mouth Stimulation of the sympathetic, however, produced a flow of saliva Division of the sympathetic nerve or extirpation of the cervical ganglion did not evoke a paralytic secretion Rawlinson (1935) repeated these observations in Babkin's laboratory and found, on examination of the gland, that there was evidence of secretory activity during the paralytic secretion in the serous cells Moreover, these cells were more excitable than normal on stimulation by adrenalin or on electrical stimulation of the sympathetic nerves Evidently, cutting the chorda tympani causes some alteration in the gland acini which greatly increases the excitability of the sympathetic nerve endings on the serous or demilune cells Emmelin (1952) points out that only under conditions where a pronounced supersensitivity has been established by denervation and an increase in release of epinephrine occurs does paralytic secretion exist Removal of the adrenal glands from the animal or the injection of sympatholytic agents such as dihydroergotamine abolishes paralytic secretion Thus it appears that paralytic secretion is due to epinephrine secreted by the animal's suprarenal glands, and that the degree of response will depend upon the level of sensitivity of the denervated salivary gland Atropine has no effect on paralytic secretion

Salivary Response to Electrical Stimuli

The importance of neural mechanisms in the control of salivary secretion has long been recognized, thus, experimentation by applying electrical stimulation to the parasympathetic and sympathetic nerves supplying the salivary glands has contributed greatly to our knowledge of salivary secretion Histological and electrical changes can be produced in the acinar cells by electrical stimulation In addition, stimulation of the salivary nerves may affect structures other than the secreting cells and thus modify secretion For instance, marked vascular changes within the gland can result from electrical stimulation In addition, electrical stimuli can presumably cause contraction of the myo-epithelial cells and produce a temporary acceleration of the flow of saliva being secreted

Electrical stimulation of the nerves to the salivary gland produces different results in different species of animals Stimulation of the parasympathetic nerves usually produces an abundant flow of quite watery saliva regardless of the gland or species studied Stimulation of the sympathetic secretory fibers to the submaxillary or sublingual glands, although causing an abundant secretion in the cat, causes only a scanty and very viscid secretion in the dog

PARASYMPATHETIC STIMULATION

Stimulation of the parasympathetic nerve supply to the salivary glands produces a prompt and abundant flow of watery saliva from the gland Vascular vasodilation within the gland also occurs The secretory effect of parasympathetic stimulation was discovered by Ludwig (1851) when he observed a flow of saliva from the submaxillary gland on stimulation of the chorda tympani nerve The vasodilator effect, associated with chorda tympani stimulation, was described by Claude Bernard in 1858 Most workers today consider the vasodilation as secondary to the process of secretion and due to the release of vasodilator agents from the activated secretory cells

The marked secretory effect that can be obtained on parasympathetic stimulation is illustrated by the fact that a salivary gland can secrete an amount of saliva equal to its own weight in just a few minutes With repetitive parasympathetic stimulation, the rate of salivary secretion within certain limits is proportional to the rate of stimuli being applied In the cat, the maximal secretory response from the submaxillary gland is obtained with a rate of 10 to 12 stimuli/sec When the maximal rate of secretion has been attained by parasympathetic stimulation, this rate cannot be exceeded by the addition of other stimuli as epinephrine or acetylcholine or even by stimulation of the sympathetic nerve supply to the gland

Lundberg (1955) has estimated that each acinar cell is supplied with from 5 to 10 parasympathetic fibers By histochemical techniques, Snell and Garrett (1956) have demonstrated true cholinesterase in the fine network of neural fibers around the acinar cells

SYMPATHETIC STIMULATION

Stimulation of the sympathetic nerve supply to the salivary glands produces a much smaller and more variable response than stimulation of the parasympathetic In the cat, sympathetic stimulation usually produces an increase in secretion from the submaxillary gland and little effect is noted upon the parotid gland In the rabbit, on the other hand, the response from sympathetic stimulation is more marked from the parotid gland than the submaxillary gland As early as 1851, Ludwig observed a flow of saliva from a submaxillary gland on stimulation of the cervical sympathetic trunk Claude Bernard in 1858 called attention to the fact that marked vasoconstriction also occurred with stimulation of the cervical sympathetic trunk

The marked vasoconstriction that occurs on sympathetic stimulation is capable in itself of altering salivary secretion With marked vasoconstriction, the blood supply to the gland may be restricted to such an extent that secretion may be inhibited

ELECTRICAL PHENOMENA ASSOCIATED WITH SALIVARY SECRETION

Stimulation of the autonomic nerves to the salivary glands causes characteristic changes in the electrical potentials that may be observed by placing electrodes in various positions on or within the gland, and connecting them to a suitable recording device With one electrode on the hilum and another on the outer surface of the submaxillary gland, stimulation of either the sympathetic or parasympathetic nerves causes the hilum of the gland to become elec-

tropositive to the outer surface after a comparatively long latent period The latent period is 0 2 to 0 3 sec with the parasympathetic, and somewhat longer when the sympathetic is stimulated Following parasympathetic stimulation, the hilum positivity rapidly declines and may change to a negative potential, there is another sharp rise in hilum positivity when the current is turned off The potential caused by sympathetic stimulation is more consistently hilum-positive

Since the submaxillary gland is composed of several different types of cells, each of which may have its own electrical characteristics, it is doubtful what significance should be ascribed to the external secretory potential More precise information may be obtained from the use of micro-electrodes with the tip placed in the interior of a single cell With these, Lundberg (1957) has obtained three types of response representing presumably three types of cells mucous cells, serous cells, and duct cells In all cases, the interior of the cell was electro-negative to the external medium In the case of the mucous cells, stimulation of either the parasympathetic or sympathetic caused a sustained increase in internal negativity or "hyperpolarization" of the cell membrane This followed a time course roughly parallel to that already described for the external secretory potential When the micro-electrode was in what was believed to be a serous cell, the parasympathetic caused hypopolarization, but the sympathetic caused a decrease in the internal negativity of the cell (apparent depolarization) With the electrode in what was thought to be a cell of a salivary duct both sympathetic and parasympathetic nerves caused apparent depolarization

In the sublingual gland, micro-electrodes picked up currents corresponding to those obtained from acinar cells of the submaxillary, but the external secretory potential was of a polarity opposite to that of the submaxillary gland, that is, the hilus became electronegative This difference was believed to be due to differences in the cellular composition of the two glands and not to any difference in the electrical response of corresponding individual cells

These electrical phenomena are remarkable in several respects Evidently they are not analogous to the familiar action potentials of nerve and muscle. They appear not to be concerned with the excitatory process (which must be quite different in gland cells and in muscle cells) but are concerned with the functional work of the cell Lundberg (1958) has presented good evidence that the hyperpolarization, associated with secretory activity, is caused by an active transport of chloride ions through the outer cell membrane which is triggered by the release of a neurohormone at the autonomic nerve endings The apparent depolarization, when it occurs, is not a self propagating disturbance of the cell membrane of the sort we are accustomed to associate with depolarization, but is probably due to the sustained activity of an ion transport mechanism in the duct cells opposite to that causing hyperpolarization in the acinar cells

As a corollary to these observations, Lundberg (1958) has suggested that secretion of water and salt by the acinar cells is accomplished through the activity of an active chloride ion transport mechanism in the outer cell membrane which "pumps" chloride ions into the cell from the surrounding interstitial fluid Electrical forces cause sodium to accompany chloride, and osmotic forces take in water As a result of the increased hydrostatic pressure within the cell, the salt and water escape through the inner membrane (where there may also be ion pumps) into the lumen of the acinus He has suggested further that the rodded epithelial cells of the ducts actively absorb sodium from the secretion Sodium absorption at this point would account for the apparent depolarization of these cells (decreased internal negativity) and also for the fact that saliva may be hypotonic with respect to the blood

Functions of Saliva

Saliva has many functions, but its main functions are to aid in digestion and to provide a protective secretion in the mouth that keeps the mucous membranes moist and aids in oral hygiene Saliva is not essential to life, but its absence results in a number of inconveniences The salivary glands, by secreting saliva, aid in the recognition and monitoring of thirst and body water needs

PREPARATION OF FOOD FOR SWALLOWING

Lubrication is perhaps the most important function of saliva When food is introduced

into the mouth, salivary secretion is increased The saliva thus provided serves to moisten the food and the mucous membrane and greatly facilitates chewing and swallowing

SOLVENT AND CLEANSING ACTION

Taste is chemically mediated, therefore, as a prerequisite to stimulation of the taste buds, any substance must be dissolved It follows that deprivation of saliva would greatly limit taste sensibility It has been mentioned previously that saliva contains little or no glucose and that the concentration of sodium is considerably below that of plasma This is probably significant in the perception of the taste of sweet or salty substances

The constant flow of saliva exerts a very necessary cleansing effect on the mouth and teeth, which are rinsed and kept comparatively free from food residues, shed epithelial cells, foreign particles, etc In this way, saliva inhibits the growth of bacteria by removing the material that may serve as culture media In addition, lysozyme in saliva serves a protective function through its ability to lyse certain bacteria One has but to consider the foul condition of the mouth in certain fevers when salivary secretion is suppressed to realize its importance in oral hygiene When salivation is suppressed, decomposing organic materials swarming with bacteria (sordes), tend to coat the teeth and lips and must be removed by artificial means The moistening of the mucous membranes of the mouth and lips by saliva also aids in articulation When the supply of saliva is insufficient, public speakers must sip water frequently in order to counteract the desiccating effect of evaporation from the mouth

DIGESTIVE FUNCTIONS

The principle substrates in the diet that amylase acts on are starch and glycogen On complete digestion, amylase reduces starch to approximately 88% maltose and 12% glucose The site of action of amylase is at the α,1-4, glycosidic linkage of starch or other oligosaccharides

Little digestion of dietary carbohydrates occurs in the mouth, pharynx and esophagus It has been shown that amylase activity continues in the stomach and that 60 to 70% of a starch or potato meal will be digested in the stomach before the meal is emptied into the duodenum Pancreatic amylase continues the digestion of oligosaccharides in the small bowel

EXCRETORY FUNCTION

Many substances, both organic and inorganic, are excreted in the saliva Mercury, potassium iodide and lead, when introduced into the body, are excreted in part in the saliva Severe inflammation of the oral mucous membrane may develop as a result of the excessive excretion of mercury via the saliva After entering the mouth via the saliva of persons with lead poisoning, lead sulfide is deposited in the gums and appears as a blue line Sulfur derived from organic materials is deposited in the tartar on the lingual side of the teeth near the gum line This explains why discoloration of the gums does not occur where teeth are absent In patients with chronic nephritis, the saliva contains a high percentage of urea Alkaloids as morphine, and antibiotics as penicillin, streptomycin, and aureomycin are also excreted in saliva

Several types of microorganisms, some intensely virulent (e g , the viruses of rabies and poliomyelitis) are excreted in the saliva The latter disease has been reproduced in monkeys by injecting the saliva of an infected person In this connection, it may be added that mumps, which is usually looked upon as a specific inflammation of the parotid gland, is in reality a systemic disease since other organs, including the ovary or testicle, breast, pancreas, and meninges, are often also inflamed The parotid may become inflamed secondarily as a result of passage of the infectious agent through the gland into the saliva

In spite of their versatility in regard to excretory functions, this role of the salivary glands is not an indispensable one The kidney represents the major organ of excretion Doubtless all the excretory functions of the body could be carried on effectively in the absence of the excretory function of salivary glands

Effects of Chemicals and Drugs upon Salivary Secretion

Many chemicals and drugs are capable of affecting salivary secretion Drugs that mimic sympathetic or parasympathetic nerve stimulation result in an increase in salivary secretion *Sympathomimetic* agents that stimulate salivary secretion include epinephrine, norepinephrine, synephrine and amphetamine *Parasympathomimetic* agents causing salivary secretion include acetylcholine, methancholine (mecholyl) and pilocarpine Cholinesterase in-

hibitors known as *anticholinesterases* provoke salivary secretion

Sympatholytic and *parasympatholytic* drugs, which block the effects of epinephrine and acetylcholine respectively, produce inhibition of salivary secretion Ergot, dibenamine and dehydroergotamine are examples of sympatholytic agents that inhibit secretion Parasympatholytic or *anticholinergic* agents such as atropine, scopolamine, and methylscopolamine produce a marked suppression of salivary secretion Many anticholinergic agents used therapeutically (e g , in the treatment of duodenal ulcer) inhibit salivary secretion and produce dryness of the mouth as a disturbing side effect

The salivary secretion caused by various surgical anesthetics such as ether, chloroform, and cyclopropane is reflex in origin due to the anesthetic coming in contact with the buccal mucosa Because excessive salivary secretion can be troublesome to the anesthetist, atropine and scopolamine have long been used as a premedication to anesthesia to suppress salivary secretion When deep anesthesia has been induced excessive secretion stops due to central inhibition

Disturbances of Salivary Secretion

Temporary suppression of salivary secretions due to emotional states such as anxiety or fear, the presence of fever, or as a result of dehydration are not uncommon Suppression of salivary secretion may also result from obstruction of the salivary ducts due to calculi formation In patients with Bell's palsy, in which there is a peripheral paralysis of the facial nerve, secretion of the salivary glands on the paralyzed side is decreased

Permanent suppression of salivary secretion (*xerostomia* or *aptyalism*) is rare and may be associated with congenital hypoplasia or absence of the salivary glands Steggerda (1941) has reported an interesting case of a university student in whom the orifices of the ducts of the salivary glands were not demonstrable The chronic dryness of the subject's mouth was not necessarily related to thirst The student had learned to take a few swallows of water approximately every hour to relieve his uncomfortable xerostomia When he became "actually thirsty" he would drink about 250 ml of water This occurred approximately four times a day He found that it was necessary to take small sips of water frequently during his meals to facilitate mastication and swallowing The water intake of this subject was studied for a number of days and found not to be different from that of four students who served as controls

Hyposalivation may also result from surgical procedures or x-ray therapy in areas where the principal salivary glands are located In xerostomia, the patient complains of difficulty in masticating and swallowing food Articulate speech is also difficult unless the subject takes frequent sips of water In such patients, there is also an increase in dental caries

Hypersalivation, or *sialorrhea*, is sometimes observed during pregnancy The cause of the excessive salivation is not known It may be particularly troublesome because the patient may secrete as much as 3 to 4 liters of saliva per day Neoplasms of the mouth or tongue, or even a carious tooth, may result in hypersalivation due to reflex stimulation of the salivary glands due to local irritation Excessive salivation is not uncommon in diseases of the esophagus, stomach and pancreas Neoplasms of the esophagus, or even esophageal spasm, may result in reflex stimulation of salivary secretion (esophago-salivary reflex) This reflex may be elicited readily in a normal person by the passage of a stomach tube or an esophageal sound Patients with gastric or duodenal ulcer and pancreatitis may also have excessive salivary secretion

Various neurological disorders, such as the parkinsonian syndrome, are associated with excessive salivation The psychotic state of schizophrenia may also be associated with excessive salivation

An interesting disorder of salivary secretion is known as the *chorda tympani syndrome* This syndrome results from a surgical or traumatic severance of the chorda tympani nerve The regenerating chorda nerve fibers become misdirected and grow into nerves innervating sweat glands in the submental region When the subject eats, or when salivary secretion is otherwise stimulated reflexly, sweating occurs in the submental area on the affected side Both the salivary glands and the sweat glands are innervated by cholinergic fibers In the chorda tympani syndrome, impulses intended for the salivary glands are transmitted to the sweat glands and they respond by secreting sweat.

5 / GASTRIC SECRETION

The stomach has the amazing capacity to secrete hydrochloric acid at a concentration of 160 mEq /liter, the proteolytic enzyme pepsin, the intrinsic factor needed for vitamin B-12 absorption by the ileum, the polypeptide hormone gastrin which stimulates the parietal cell to secrete the highly acid juice already listed, as well as a variety of mucoid substances including blood group substances

Structure of the Gastric Mucosa

The surface epithelium of the gastric mucosa consists of a single layer of tall columnar cells that secrete mucus The mucous membrane of the stomach is separated from the outer muscular gastric wall by the *muscularis mucosae* and a supporting stroma of connecting tissue The surface of the gastric mucosa is marked by slight linear depressions into small polygonal areas called *areae gastricae*, giving the surface a mammillated appearance, the areas measure from 1 to 6 mm in diameter The surface is constantly covered by a thick layer of tough, tenacious mucus secreted by the surface epithelial cells The layer of mucus varies in thickness from 0 5 to 2 5 mm Beneath the surface layer of epithelium are located various types of tubular glands which vary in structure and in composition of their secretion in different parts of the stomach When the mucus is removed, the *gastric foveolae* may be seen with the aid of a hand lens The gastric foveolae are separated from one another by a distance of approximately 0 1 mm , thus, there are 90 to 100/mm ^{2} Opening into the bottom of each foveola are three to seven individual gastric glands The lining of the gastric pit is a continuation of the tall columnar surface epithelium

The gastric mucosa can be divided into several areas, each of which contains a particular type of gland The *cardiac area* is the zone, 1 to 4 cm wide, that guards the esophageal orifice The *fundic area* is the largest area of the stomach, accounting for 60 to 80% of total mucosal surface, interposed between the cardiac and the *pyloric* areas The lower limit of the fundic area is marked by the incisura angularis on the lesser curvature of the stomach, and a line extending diagonally toward the pylorus to the greater curvature The junction of the pyloric and fundic area is not sharply demarcated and is frequently referred to as the *transitional zone* The distal part of the pyloric zone extends to the pylorus which communicates with the first part of the duodenum The pyloric area is about 15% of the total gastric mucosal area On the lesser curvature, where the mucous membrane is more firmly attached to the muscular wall of the stomach, the rugae form parallel longitudinal folds and this area is sometimes referred to as the *Magenstrasse*

The gastric glands are tubular in structure and extend from the bottom of the gastric foveolae to the muscularis mucosae where they terminate in a blind bulbar end known as the *fundus* of the gland The main tubular portion of the gland is known as the *body* The *neck* of the gland connects the body to the *isthmus* which communicates with the gastric crypt The total number of gastric glands has been estimated at 35 million

Each area of the gastric mucosa is characterized by its glandular structure The *cardiac glands* tend to be short and tortuous and are made up mostly of mucous cells, although a few pepsinogen cells are also present Similar glands, minus the pepsinogen cells, are found in the esophagus The *fundic glands* tend to be straight, slender structures with narrow lumen Three types of cells make up the fundic glands *mucous* cells, *pepsinogen* or *chief* cells, and *parietal* or oxyntic cells The mucous cells may be further divided into isthmus cells and neck cells Although the isthmus cells are of the mucous type, they probably secrete very little mucus The cells of the neck are true mucous-secreting cells but they secrete a different type of mucus from that of the surface epithelial cells The mucus of the neck cells is soluble, and constitutes a part of the soluble mucus of gastric juice An important characteristic of the isthmus and neck cells is their frequent mitoses It is believed that these cells may develop into pepsinogen cells and thus replace chief cells

In the neck region are also a great many par-

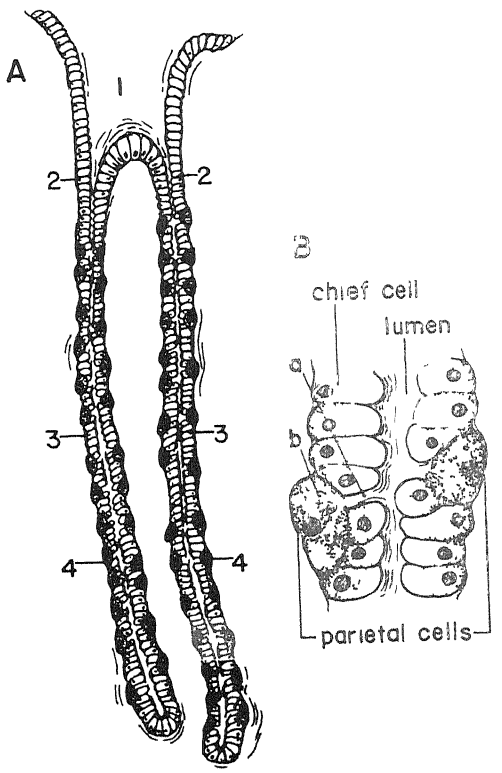


Fig 2 13 *A* gland from the fundus of the stomach *1* crypt or foveola on the mucous surface connected to the neck (*2*) by the isthmus *2* neck of gland containing mucin-secreting cells *3* chief cells *4* parietal cells *B* enlarged drawing showing (a) intercellular and (b) intracellular canaliculi

ietal cells These cells do not form a part of the gland wall but appear to lie outside, giving the gland a nodular appearance (Fig 2 13) The parietal cells secrete the hydrochloric acid and most of the water of the gastric juice and, because of this, they are also known as the "oxyntic" or acid-secreting cells Although separated from the lumen of the tubules by the chief, or pepsinogen cells, the parietal cells transmit their secretion into the lumen through delicate canaliculi lying between the chief cells These canaliculi are extensions of a system of exquisitely fine canals which lie within the protoplasm of the parietal cells. Numerous transparent, smooth-surfaced vesicles are seen in the protoplasm of the cells during rest Depletion of this vesicular material occurs during activity The distribution of parietal cells in the human stomach is shown in Figure 2 14

The body of the gland which makes up the remainder of the tubule is composed principally of pepsinogen cells called *chief cells* Parietal cells are also present, but in diminishing numbers, as the fundus of the tubule is approached In addition to *pepsinogen*, the chief cells probably produce *renin* (in the calf), which coagulates milk, and is believed to be produced only in young animals, and a special gelatin-splitting enzyme known as *gelatinase* (in the pig), which liquifies gelatin at a rate about 400 times that of pepsin It is not known whether each of these enzymes is produced by a different type of chief cell or whether they are produced by the same cell Since the chief cells appear to be morphologically identical, there is no anatomical evidence for assuming that they differ from one another in their secretory products No mitoses are seen in the chief cells

Another type of cell occasionally found in the fundic glands is the argentaffin cell The granules contained in their protoplasm are easily impregnated with silver and this characteristic gives the cells their name The exact function of argentaffin cells in the gastric mucosa is not known

The *pyloric glands* are located in the pyloric area but they also extend into the transitional zone separating the pyloric area from the fundic area of the stomach The pyloric glands tend to be somewhat shorter and are more tortuous than the fundic glands The foveolae in the pyloric region tend to be deep These glands are lined exclusively by mucous-secreting cells that resemble those of the cardiac

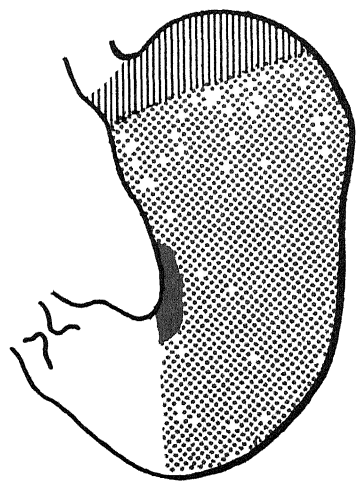


Fig 2 14 Diagram showing the distribution of the parietal (acid-secreting) cells in the human stomach In the black area the proportion of parietal cells is maximal and is taken as 100% in the brown area on lesser curvature the percentage of parietal cells is 75% in the shaded area at the fundus, 50% and in the white area 0 to 1% (After Berger 1934)

glands An occasional argentaffin cell may be noted in the fundus of the pyloric glands Gastrin has recently been found to be contained in or on abundant granules packing the cytoplasm of differentiated cells interspersed in the antral glands These cells appear endocrine-like, and do not take the silver stains

The electron microscope has provided additional information on the structure of the gastric mucosa The presence of intracellular canaliculi in the parietal cells is a constant finding At the free or luminal surface of the parietal cell, as in the case of certain other active transport cells, there is a striated border The intracellular canaliculus simply represents an invagination of this striated border The presence of a very high concentration of mitochondria also is consistent with the cell's role in transport of ions against a concentration gradient (Fig 2 15)

Studies correlating structural change with gastric secretion (Sedar and Friedman, 1961,

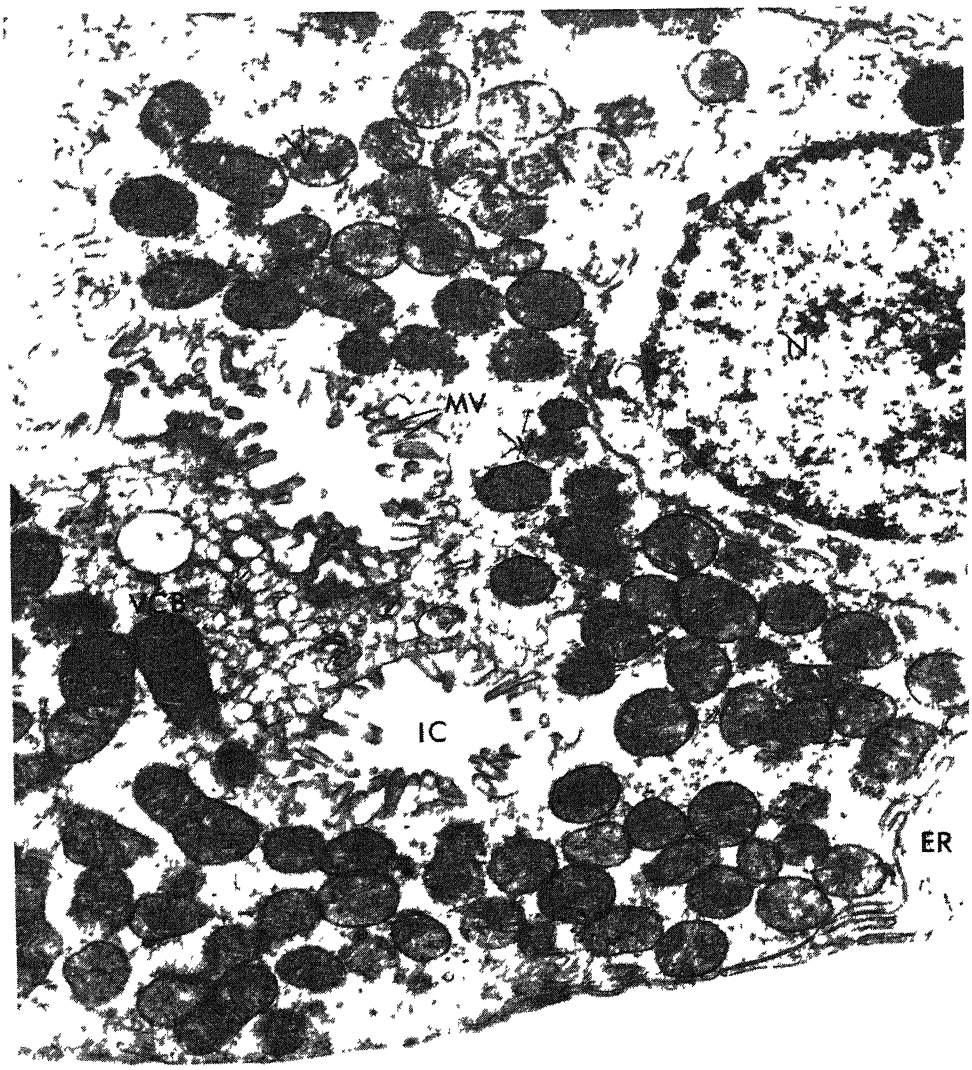


Fig 2 15 Electromicrograph of part of a nonsecreting human parietal cell The specimen was obtained by biopsy Secretion was suppressed by administering 0 5 mg of atropine sulfate The gastric juice contained no free acid and had a pH of 6 6 at the time of biopsy The intracellular canal system (IC) is evident demonstrating patent lumen with projecting microvilli (MV) Throughout the cytoplasm are many small vesicular structures (V) whose content appears less dense than that of surrounding cytoplasm Note the large vacuole containing body (VCB) Nucleus (N) and some endoplasmic reticulum (ER) of adjacent chief cells are also included in electromicrograph (Approx × 22 800) From Rohrer et al 1965)

Sedar, 1962, Rohrer et al , 1965) have demonstrated that during gastric secretion stimulated by histamine, insulin hypoglycemia, or vagal stimulation, definite and characteristic changes occur within the parietal cell (Fig 2 16) The microvilli lining the walls of the canaliculi are markedly increased, to the point that the lumen of the canaliculus is almost occluded by these protoplasmic projections The increased number of microvilli provides a tremendous increase in the surface area of the wall of the canaliculus, which is presumed to be the site of movement of hydrochloric acid and water from the cell into the canaliculus During secretion of hydrochloric acid, the smooth-surfaced vesicles, which are randomly distributed in the cytoplasmic matrix in the resting cell, become concentrated adjacent to intracellular canaliculi With continued secretion of hydrochloric acid, these smooth-surfaced elements decrease in number In some instances it has been possible to observe these vesicles communicating with areas of the canalicular wall between adjacent microvilli



Fig 2 16 Part of parietal cell from same subject as in Figure 2 15 Biopsy was taken after maxima histamine stimulation when gastric juice contained 104 mEq /liter HCl and had pH of 1 Intracellular canaliculus (IC) appears very complex and lumen is filled with microvilli (compare with Figure 2 15) Vesicular components (V) appear decreased in number and are oriented about intracellular canal In cytoplasm are mitochondria (M) a few scattered tubular elements (TE), ribonucleoprotein granules (RNP), and two dense bodies (DB) An infolded cell membrane (CM) is noted on right (From Rohrer et al 1965)

Innervation of the Gastric Glands

The gastric mucosa is innervated by both the sympathetic and parasympathetic divisions of the autonomic nervous system The *efferent sympathetic supply* arises from neurons located in the latter horn of the 5th to 10th thoracic segments of spinal cord The *preganglionic fibers* emerge from the cord in the anterior spinal nerve roots and pass to the ganglia of the thoracic sympathetic chain Here, some form synapses but most continue in the thoracic splanchnic nerves to the celiac plexus and form synapses with the ganglia within the celiac plexus From the celiac ganglia, the *postganglionic fibers* accompany the arterial vascular supply of the stomach *Afferent visceral fibers* from the stomach travel the same course just outlined, in reverse, to the ganglion cells in the posterior spinal nerve roots The afferents do not synapse in, or arise from, the sympathetic ganglia

The *efferent parasympathetic supply* to the gastric mucosa originates in the dorsal vagal nucleus in the floor of the fourth ventricle The *preganglionic fibers* reach the stomach via the right and left vagus nerves The left vagus supplies preganglionic fibers to the anterior surface, and the right vagus to the posterior surface of the stomach In addition, both vagi send branches to the celiac plexus The parasympathetic preganglionic fibers accompany the vascular supply and the sympathetic fibers into the muscular layers of the wall of the stomach, to form synapses with the ganglion cells of the myenteric plexus of Auerbach and the submucosus plexus of Meissner The *postganglionic fibers* supplying the mucosa are thus very short, whereas the preganglionic fibers are quite long *Afferent parasympathetic fibers* are found in the vagus The neurons of these fibers are located mainly in the *ganglion nodosum* The central processes of these afferent neurons terminate in the dorsal nucleus of the vagus, completing the reflex arc

The vagi are the important secretory nerves to the stomach The cephalic or psychic phase of gastric secretion is mediated via the vagi An intact vagal innervation is also important in the release of *gastrin* from the gastric antrum Stimulation of the vagus nerve causes the secretion of strongly acidic gastric juice rich in pepsin This effect is mediated through acetylcholine (Fig 2 17) Atropine, which blocks postganglionic parasympathetic stimulation,

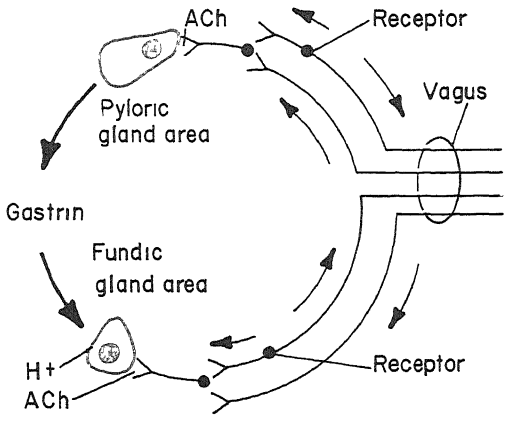


Fig 2 17 Diagrammatic representation of the intramural and vagal innervation of gastric mucosa (From Grossman 1963)

inhibits gastric secretion stimulated via the vagi It should be emphasized that the vagi contain thousands of afferent fibers which probably mediate reflexes originating in the gastrointestinal tract

Stimulation of the sympathetic nerve supply to the mucosa causes the secretion, mainly from the pyloric glands, of an alkaline mucoid juice which is very low in peptic activity The sympathetic response is unaffected by atropine, but is annulled by ergotamine (which paralyzes motor and secretory sympathetic fibers) Vineberg's results (1931) indicate that the vagus controls the secretion of mucin by the mucous neck cells and the surface epithelium of the gastric mucosa The influence exerted by the sympathetic upon the peptic and oxyntic cells of the gastric body is not definitely known According to some, its effect upon these elements is inhibitory

Origin and Composition of Gastric Juice

Gastric juice is the composite product of secretion by the surface epithelium and the various gastric glands The composition of gastric juice varies, depending upon the individual contributions of the various secreting structures The major constituents of gastric juice are water, hydrochloric acid, and other electrolytes as sodium, potassium, calcium, phosphate, bicarbonate, and sulfate The major organic constituents include various enzymes, at least two types of mucus, the *intrinsic factor* of Castle, and blood group substances

It has been estimated that the volume of gastric juice secreted by man in a 24-hr period amounts to 1200 to 1500 ml The amount will vary, depending upon the diet and other stimuli provoking gastric secretion When gastric juice is collected for a 12-hr period during the night, by means of a nasogastric tube with constant suction, the average volume obtained is about 400 ml Fasting human gastric juice collected from healthy young adults has an acidity of 40 to 60 mEq /liter, a pH of 0 9 to 1 2, and a specific gravity of 1 002 to 1 004 Under histamine, 1 ml of mucosa in man can produce 1 mEq of HCl/hr

HYDROCHLORIC ACID AND GASTRIC ELECTROLYTES

The chief problem is to account for the variation in acidity and electrolyte composition of gastric juice based on the main facts parietal cells secrete water and HCl, with a maximal concentration about 160 mEq /liter The main anion is chloride, and when the acidity is low, the other cation is sodium Na^+ and K^+ are always present At times, HCO_3^- up to 25 mN can be detected It is convenient to think of gastric juice as a mixture of parietal and nonparietal secretions

The composition of pure parietal secretion can be arrived at only by computation, inasmuch as no means is available for collecting pure parietal secretion that is not contaminated, at least to some extent, with secretions from nonparietal cells Figure 2 18 illustrates the computed ionic composition of parietal and nonparietal secretions of the gastric glands It is further believed that the composition of parietal secretion remains constant, and is independent of rate of secretion, nature of stimulus, or strength of stimulus The concentration of hydrochloric acid in gastric juice collected from man or a dog may vary from 0 to 150 mEq or more/liter Other things being equal, the concentration of acid increases with increasing rates of secretion, up to a maximum of about 150 mEq /liter The chloride ions that accompany the hydrogen ions into the gastric juice to form HCl are not passively carried along by the hydrogen ions Two important observations that indicate that the chloride ion of the HCl in gastric juice is secreted by an active, energy-consuming mechanism are *1)* chloride ion is transported against a concentration gradient, from a plasma concentration of about 110

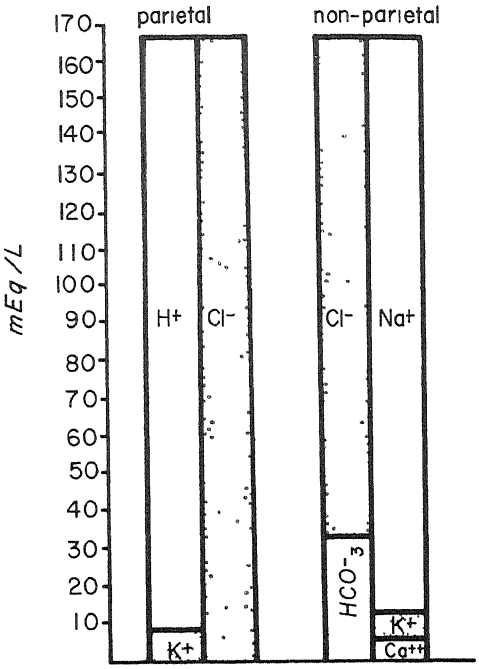


Fig 2 18 Computed ionic concentration of the principal ions in the parietal and nonparietal secretions of the gastric gland (Redrawn from Ivy et al 1950)

mEq /liter, to a concentration in gastric juice up to 166 mEq /liter, and 2) chloride ion is transported against an electrochemical gradient (the luminal surface of the gastric mucosa is negative to the serosal surface with a potential difference of 30 to 50 mV) The chloride ion of HCl is considered to be secreted by a "chloride pump "

It is interesting that the mechanism (chloride pump) for transporting chloride ions into the gastric juice is also capable of transporting *bromide* and *iodide* ions Hydrobromic acid may be recovered from the gastric juice of patients who have taken bromide salts When radioactive iodine, ^{131}I, is administered intravenously to a patient, radioactivity rapidly appears in the gastric juice

A small part of the chloride of gastric juice is balanced against metallic cations, principally sodium This constitutes the so-called *neutral chloride* The concentration of other negative ions in the juice is negligible, thus the osmotic activity of gastric juice is primarily due to the acid, plus the neutral chloride Since gastric juice is practically isotonic with blood, the ionic concentration of total chloride in the gastric juice differs but little from that of the total base of the blood

In highly acid gastric juice, the neutral chloride is present only in small amounts, but in a less acid secretion it may constitute a major portion of the total chloride When it is present in large quantities, its source becomes a matter of considerable interest The source of neutral chloride has been a matter of widespread discussion and controversy but current thinking tends to the view that it is secreted by the nonparietal cells (mucous cells of the surface epithelium and chief cells of the gastric glands), along with sodium, calcium, potassium, and the organic materials secreted by these cells, and that a certain, perhaps smaller, amount enters the gastric contents by diffusion from the interstitial fluid In impure gastric juice, a large part of the neutral chloride may represent swallowed saliva and regurgitated duodenal contents

GASTRIC ENZYMES

Pepsin. The principal proteolytic substance in gastric juice is pepsin, from the chief cells of the fundic glands The zymogen in the cytoplasm of these cells is inactive and referred to as *pepsinogen* Pepsinogen is relatively resistant to heat and alkali and can withstand short exposures to temperatures of 70° C or a pH of 11 without loss of activity The enzyme becomes active only in acid media At a pH below 6, pepsinogen is transformed into pepsin by an autocatalytic reaction The optimum pH for peptic activity is about 2 0 At a pH of 5 0, the action of pepsin is almost abolished Pepsin digests protein by attacking those peptide linkages in which the amino groups are attached to aromatic amino acids The products consist mainly of proteoses and peptones and relatively few amino acids and polypeptides Pepsin is constantly present in acid gastric juice and probably is secreted continuously During interdigestive periods, it is stored in the chief cells of the fundic glands, evidence by the accumulation of zymogen granules Details of the synthesis of pepsinogen from amino acids by the chief cells, or the mechanism of release of pepsinogen from the cells during secretion are not known Samloff and Townes (1969) have demonstrated the existence of seven electrophoretically distinct pepsinogens in human gastric mucosa Five are limited to the fundus and body, two are in the antrum and proximal duodenum This pattern is under genetic control, with some families lacking one specific pepsinogen The major stimulus to pepsin secretion is vagal impulses Gastric secretion induced by sham feeding, electrical stimulation of the vagus nerves, or insulin hypoglycemia results in a gastric juice containing high peptic activity Pilocarpine and acetylcholine are potent stimuli of pepsin secretion Now, however, we know that gastrin itself can stimulate pepsin secretion, independent of vagal stimulation Histamine produces a gastric juice of low pepsin concentration, but does stimulate pepsin in man

Most stimuli of gastric secretion produce a juice that is initially high in peptic activity Gastric secretion in response to a meal will show the highest concentration of pepsin in the 1st hr As the volume of gastric juice increases, the concentration of pepsin falls During the later stages of digestion, when the volume of gastric juice diminishes, the concentration of pepsin increases Since pepsin is a preformed secretory product and cannot be elaborated as rapidly as it is secreted, depletion of cell stores of pepsinogen occurs during the initial period of gastric secretion Prolonged stimulation of gastric secretion, however, does not result in disappearance of peptic activity from gastric juice

Proteolytic activity in blood serum and urine, presumably due to the presence of pepsinogen, has been demonstrated Apparently, a small amount of pepsinogen leaks into the blood Pepsinogen can withstand the slightly alkaline pH of blood and its molecular size (42,500) would permit it to be filtered by the renal glomerulus The name *uropepsin* has been given to the protease in urine

Rennin. This enzyme is the milk-curdling ferment Rennin has been isolated from the gastric secretions of the calf The optimum pH for the action of rennin is between 6 0 and 6 5 The presence of calcium ions are necessary for the action of rennin on milk Gastric rennin, when added to milk kept at body temperature, causes a change in the soluble *casein*, splitting it into a *proteose-like* substance (whey protein) which remains in solution, and *paracasein* The latter is soluble alone but upon combination with calcium forms an insoluble compound, *calcium paracasein*, which precipitates as a curd The protein (paracasein) of the curd subsequently undergoes peptic digestion in the usual manner The softer and more finely flocculent the character of the curd that is formed,

the more readily is it attacked by the proteolytic ferments

Rennin is not found in the gastric juice of man either infants or adults The clotting and digestion of milk in man is by pepsin With an optimum pH of 6 0 to 6 5 rennin would be inactive at the usual pH of the gastric contents of the normal adult

Lipase A gastric lipase has been found in gastric juice The enzyme is a tributerase and may be similar to plasma lipase which is also a tributerase The optimum pH for the action of gastric lipase is between 4 0 and 5 0 and is inactive at a pH of 2 5 The origin of lipase found in gastric juice in the rat is the fundic glands When compared to pancreatic lipase, gastric lipase is a weak fat-splitting enzyme

Other Gastric Enzymes. A number of less important digestive enzymes have been detected in gastric juice Their presence in gastric juice in minute amounts suggests that they are of doubtful digestive significance

Lysozyme (pH optimum 5 3) is a carbohydrate-splitting enzyme In gastric juice from normal individuals, it is found in small amounts Its cellular origin is unknown The lysozyme content of gastric juice has been found to be increased in patients with gastric ulcer This increase in lysozyme concentration is probably secondary to the inflammatory reaction produced by the ulcer

Urease can be isolated from the gastric mucosa and gastric juice Urease is found also in bacteria contaminating the gastric mucosa The hydrolysis of urea by urease to produce ammonia probably plays a negligible role in gastric secretion

Carbonic anhydrase is found in gastric juice in small amounts and its presence is thought to be due to disintegration of desquamated surface epithelial cells In addition to being present in the surface epithelial cells, carbonic anhydrase is found in high concentrations in the parietal cells, where it plays an important role in the process of HCl formation

GASTRIC MUCUS

Two types of gastric mucus occur in gastric juice *Visible* mucus is secreted by the surface epithelial cells of the gastric mucosa and is thick, viscous, and even jelly-like It forms a transparent coating over the surface of the gastric mucosa 2 to 3 mm in thickness Mucus secretion from the surface epithelium is stimulated by various chemicals applied to the surface of the gastric mucosa e g alcohol, ether, clove oil and phenol red In addition, tactile stimuli such as rubbing the mucosa with a brush, or even roughage in the diet, cause secretion of mucus by the surface epithelial cells

Soluble, or transparent, mucus is secreted by the cells of the pyloric and cardiac glands, and also by the mucous neck cells of the fundic glands It is therefore a constituent of pure gastric juice (dissolved mucin) Vagal stimulation evokes a secretion of mucin from the mucin secreting cells of the tubules Mucin is a complex of glycoproteins Secretion collected from pouches from the pyloric part of the stomach of dogs is mucoid, viscous, tenacious, and transparent, with a pH of 7 0 to 7 5 It coats the interior of the stomach and serves to protect the mucosa from the action of the gastric juice, but has little acid-reducing power Mucin does inhibit peptic activity This antipeptic action of mucin is due to its constituent, mucoitin-sulfuric acid, the main monomer of goblet cell mucus By virtue of its high viscosity, mucin effectively lubricates and protects the gastric mucosa from mechanical injury

Brunschwig and associates (1939) demonstrated the presence of a *gastric secretory inhibitor* that they isolated from the gastric juice of patients with pernicious anemia Blackburn and Code (1948) also demonstrated the presence of a gastric secretory inhibitor in human gastric juice and determined that the active material was in the mucin fraction of the juice The exact nature of this gastric secretory inhibitor is not known, but this inhibitor called *gastrone* has been partially purified and consists of a mucus-like substance with a molecular weight over 100,000 containing much carbohydrate, and a nondialyzable glycopolypeptide or glycoprotein with a molecular weight between 10,000 and 400,000

Mechanism of Hydrochloric Acid Secretion

The precise manner in which the gastric mucosa secretes hydrochloric acid is still unknown However, the following facts taken together point in a general way to the processes which must be involved in the secretion

1) It is generally agreed that the acid is formed by the parietal cells of the gastric glands (Davies, 1951) It is not believed to be produced in the protoplasm of these cells but to

appear first in the secretory canaliculi which communicate with the lumen of the gland

2) The acid, as secreted by the parietal cell, has a uniform concentration of approximately 0 17 M The theory of the constancy of the concentration of hydrochloric acid as secreted by the parietal cells was first put forward by Pavlov (1910) and has since been supported by many investigators, including Hollander and Cowgill (1931) This theory is mainly based on the observation that the nearer one approaches to a pure parietal secretion, the nearer the concentration of hydrochloric acid approaches 0 17 M The acidity, when plotted against the neutral chloride, gives a straight line expressing the inverse relationship between them When extrapolated, this line comes to a point representing a zero value for neutral chloride and an acidity of around 170 mEq /liter Since neutral chloride and acid chloride are inversely related, the total chloride tends to remain relatively constant However, since some of the nonparietal secretion contains relatively little chloride, the total chloride increases somewhat with increasing rate of parietal secretion and, therefore, with increasing total acidity (Figs 2 19 and 2 20)

A somewhat different opinion was expressed by Fisher and Hunt (1950), who made a study of the extensive data of Ihre (1938) on gastric secretion in human subjects They concluded that the most probable concentration of hydrochloric acid in human parietal secretion is 160 mEq /liter, and that the secretion contains, in addition, 10 mEq /liter of neutral chloride

3) The secretion of hydrochloric acid involves expenditure of energy, that is to say, no simple chemical reaction which could proceed spontaneously can explain the formation of hydrochloric acid by the gastric glands One estimate is that the secretion of 1 g -molecular equivalent of hydrochloric acid requires expenditures of 10,000 gram calories of energy

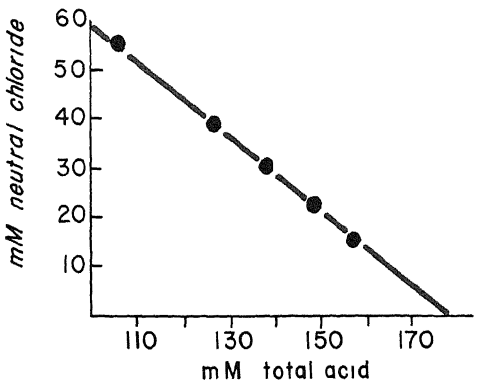


Fig 2 19 Neutral chloride as a function of acidity (After Hollander 1932a)

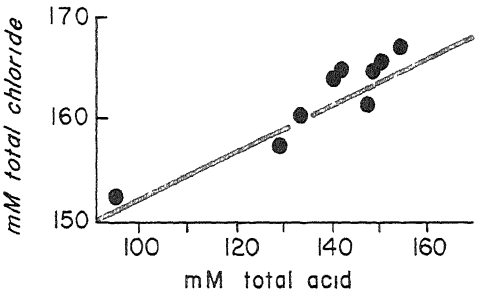


Fig 2 20 Total chloride as a function of acidity (After Hollander 1932b)

4) The energy is derived from oxidation, probably of glucose, and requires the activity of the enzymes ordinarily active in such oxidation

5) For every hydrogen ion produced, a bicarbonate ion is released into the interstitial fluid and ultimately into the blood This requires a continuous supply of carbon dioxide The carbon dioxide may be derived from the metabolism of the parietal cell itself if the secretory rate is low, but at higher rates of secretion the metabolic carbon dioxide has to be supplemented with carbon dioxide from the circulating blood The need for carbon dioxide arises from the fact that, whatever the reacting system may be, removal of hydrogen ion results in accumulation of an excess of hydroxyl ions, if these were not neutralized, the resulting alkalinity would soon destroy the secreting cells The fact that secretion of hydrochloric acid into the gastric juice results in secretion of an equivalent amount of bicarbonate into the blood explains the so-called alkaline tide seen in the urine during digestion (Bence Jones, 1850)

6) Carbonic anhydrase is present in high concentration in the parietal cells (Davenport, 1939), and participates in the overall process, though probably not directly in secretion of hydrogen ions Its apparent function is to aid in the hydration of carbon dioxide and to produce the carbonic acid that is utilized in neutralizing the hydroxyl ions produced during secretion Secretion of hydrochloric acid is greatly depressed and in some subjects, totally inhibited, by the administration of large doses of Diamox, which is an inhibitor of carbonic anhydrase (Janowitz et al , 1957) Although excised frog

gastric mucosa will secrete in the absence of carbon dioxide, the tissue is destroyed by the alkalinity resulting from the loss of hydrogen ions (Davies and Longmuir, 1948) In the living animals, at least in mammals, some mechanism apparently exists to prevent secretion under these circumstances, and this protects the secreting tissue from self-destruction

7) The overall reaction can be summarized by the equation

$$CO_2 + H_2O + NaCl \rightleftharpoons HCl + NaHCO_3$$

This equation is not intended to describe the chemical reactions in detail, but merely to indicate the beginning and end products and their quantitative relationships It indicates that the ultimate source of the hydrogen ions is water, the ultimate source of the chloride ions is the salt of the blood, for each mole of hydrochloric acid produced, a mole of sodium bicarbonate has to be formed The intermediate steps, the source or sources of energy, and the enzymes and substrates concerned in the intermediate reactions, all of which constitute the real problem of the secretion of hydrochloric acid, are not even suggested The fact that the reactions of this equation would ordinarily progress to the left is a simple qualitative proof that energy is required for the secretion of hydrochloric acid

8) Rehm (1950) and his co-workers have shown that there is a maintained electrical potential difference across the gastric mucosa, the secretory surface being negative in an external circuit to the nutrient surface When the two sides of the mucosa are connected electrically, a continuous current can be maintained representing about 10% of the metabolic energy This electrical potential decreases with the onset of acid secretion The rate of acid secretion can be increased or decreased by passing an electric current, from a battery through the mucosa so as to increase or decrease, respectively, the potential difference across the tissue Experiments with metabolic inhibitors show that most of the pathways of aerobic metabolism, and the related phosphorylations, are required to maintain the potential difference and acid secretion The facts prompted Rehm to propose a theory of acid secretion in which it is assumed that electrical energy is utilized in the separation of hydrogen ions during the secretion of hydrochloric acid It seems more likely that here, as elsewhere, the potential difference is the result of, rather than the cause of, the activity

As to the intermediate steps between carbon dioxide and water at one end of the reaction, and hydrogen ions and bicarbonate at the other end, Davies (1951), after reviewing the history of research in this field from Claude Bernard's time until 1951, had this to say "The available evidence shows that two related mechanisms of acid secretion are possible In mechanism 1, the metabolic-hydrogen atoms from glucose and water, which are transported by the dehydrogenases, become oxidized to hydrogen ions at the cytochrome level, and the electrons react with oxygen and water to form, first hydroxyl ions, and then bicarbonate ions by further reactions with carbon dioxide This process uses the oxidation reduction energy from the level of atmospheric oxygen to that of the cytochromes In mechanism 2, phosphate bond energy, generated by reactions at lower oxidation reduction levels is utilized to concentrate hydrogen ions, formed by ionization from water, in an electron cycle mechanism in which hydrogen ions are reduced to covalent hydrogen atoms, transported by a carrier system and oxidized to hydrogen ions at high concentration as a result of coupled phosphorylation Kinetic and thermodynamic considerations show that the hydrogen carrier and electron transport systems could be oxaloacetate-malate and cytochrome b, or perhaps fumerate-succinate and cytochrome c Both mechanisms require a spatial array of enzymes in the paracanalicular zone of the oxyntic cells, and in both cases, chloride ions move in the opposite direction to, and as a result of, the movements of the electrons carried by the cytochromes "

Davenport (1957) wrote 6 years later, "I suspect that the secretory process is a cyclic oxidation and reduction A carrier molecule within the cell combines with high energy phosphate The product is oxidized and the energy contained within it is dissipated in transporting the newly released proton" (hydrogen ion) "against a concentration gradient into the gastric juice Then the carrier is reduced by substrate and is ready to re-enter the cycle Two minor problems remain For each proton secreted an electron must be taken up by oxygen, and the pH of the cell must be kept down by intracellular neutralization using the carbon dioxide mechanism . " "Carbonic anhydrase which occupies a peripheral position is involved, an

enzyme behaving like succinoxidase is there, and it is likely that some sulfhydryl compound is essential" Davenport, like Davies, recognized the need for assuming appropriate orientation of the various components involved in the reaction Davenport's concept is illustrated in Figures 2 21 and 2 22

Davies (1951) considered that "the rate of transport of water by oxyntic cells is so enormous that it could not be handled molecule by molecule by any known enzyme systems The water must be moved in bulk, and probably flow osmotically as a result of the secretion of the hydrogen and chloride ions by the oxyntic cells"

Once H^+ ions have been secreted, the gastric mucosa serves to contain them and offers a barrier to the diffusion of H^+ out of the lumen, and to the diffusion of Na^+ ions into the lumen from the interstitial fluid This "mucosal barrier," as Davenport has named it, is greatest in the fundus, less restrictive in the antrum, and least in the duodenum

Control of Gastric Secretion

Gastric secretion is controlled by both neural and humoral mechanisms The parasympathetic (vagus) innervation provides the pathways for secretory stimuli to the gastric mucosa Sympathetic pathways control gastric secretion indirectly, due to their control over vasomotor mechanisms and blood flow to the mucosa

A number of humoral mechanisms also are involved in the control of gastric secretion Examples include the release of *gastrin* from the gastric antrum which *stimulates* gastric secretion, and the release of *enterogastrone* from the mucosa of the upper small bowel which *inhibit* gastric secretions The last characterized is a substance which inhibits acid secretion but stimulates pepsin output

It is convenient, when discussing the gastric secretory response to a meal, to divide the total response into different phases *cephalic*, *gastric*, and *intestinal* phases, named according to the area where the stimuli arise The volume of gastric secretion resulting from these different phases in the dog is illustrated in Figure 2 23 In man, gastric secretion tends to be continuous For this reason, a fourth phase of gastric secretion, the *interdigestive* phase, is recognized An estimate of the volume of gastric secretion of the various phases in man is shown in Table 2 7

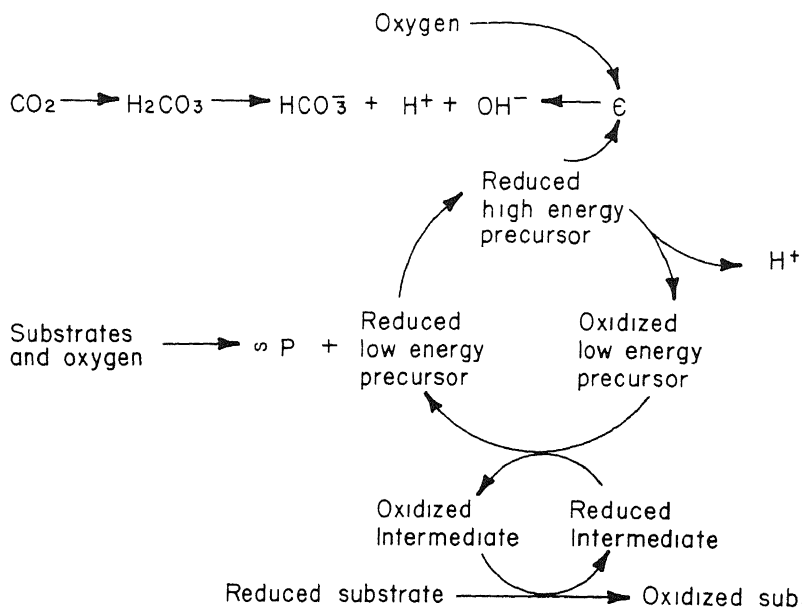


Fig 2 21 Schema of a possible mechanism of gastric acid secretion showing the four essential parts *a* generation of high energy phosphate bounds from substrate and oxygen *b* cyclic oxidation and reduction generating hydrogen ions *c* removal of electrons to oxygen and *d* intracellular neutralization by carbon dioxide (From Davenport 1957 Courtesy of University of Wisconsin Press)

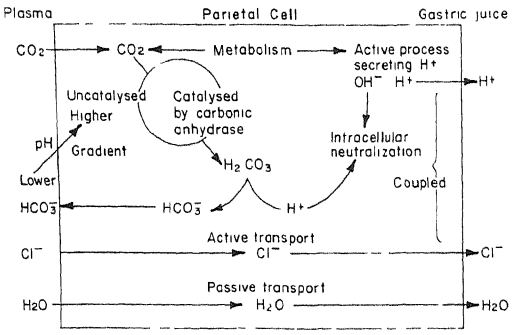


Fig 2 22 Diagram illustrating the role of carbon dioxide in intracellular neutralization during secretion of acid by the gastric mucosa (Redrawn from Davenport 1962)

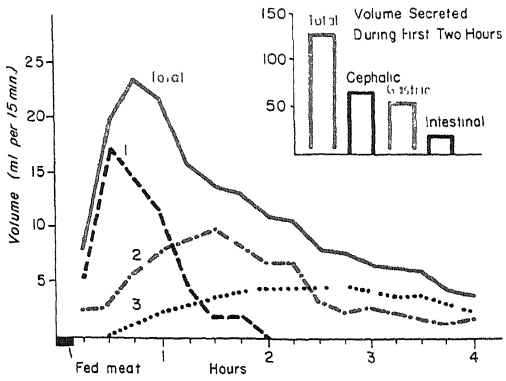


Fig 2 23 Reconstructed curves to show the approximate distribution between the several phases of gastric secretion in the dog after a meal of 300 g of meat *1* cephalic or initial reflex phase *2* gastric or pyloric phase and *3* intestinal phase (Redrawn from Thomas and Friedman 1951)

Table 2 7

Volume of different phases of gastric secretion in man

Period or Phase	Amount (ml)
Interdigestive	30–60 ml /hr
Cephalic	50–150 ml /20 min
Gastric	225–350 ml /5 hr
Intestinal	200–300 ml /5 hr

From Ivy et al, 1950, p 32

CEPHALIC PHASE OF GASTRIC SECRETION

Upon eating a meal, gastric secretion is initiated prior to the food arriving in the stomach by afferent impulses arising in the head (thus the term cephalic) which converge on the vagus nucleus, and efferent stimuli are sent to the gastric mucosa via the vagi The cephalic, or *initial reflex phase*, of gastric secretion results from the sight, smell, or taste of food, or from the act of eating and is prevented by section of the vagus nerves If the secretion results from the sight or smell of food without food making actual contact with the mouth, it is referred to as *psychic* secretion The fact that there can be an abundant secretion of gastric juice, even though there is no food in the stomach, is readily demonstrated by means of sham feeding This phase is mediated by the vagus nerves by two mechanisms *1*) direct cholinergic stimulation of the oxyntic (parietal) cells, and 2) cholinergic release of gastrin from the pyloric gland area, and mechanism 1 probably requires the presence of a small amount of gastrin itself

Janowitz and his colleagues (1950) made some interesting observations in a female subject who had a complete stenosis of the esophagus following a lye stricture and had to have a gastrostomy for feeding purposes They studied the effects of various types of meals on the cephalic phase of gastric secretion and found that an unappetizing meal consisting of cereal gruel evoked essentially no response, whereas a meal selected by the subject evoked a high secretory response The routine hospital ward diet evoked an intermediate response (Fig 2 24) The latent period of the cephalic phase response was less than 5 min The duration of the response ranged from 30 to 120 min following cessation of sham feeding and varied with the type of meal

GASTRIC PHASE

The entry of food into the stomach initiates gastric secretion by the distention from the bulk of the meal, and chemical factors in the partially digested meal alcohols, peptides, and amino acids (especially glycine, β-alanine, serine, (and lysine)

Both distention and chemical factors operate via the same two mechanisms cited for the cephalic phase *1*) direct cholinergic stimulation of oxyntic cells, and 2) cholinergic release of gastrin from the pyloric gland area cells In both, "long" (vago-vagal) or "short" local reflexes within the stomach are probably operative Figure 2 17 diagrams this

Gastrin. The most important advance in our contemporary knowledge of gastric secretion has been the purification, and structural analysis of gastrin by Gregory and Tracy (1964) This linear peptide with 17 amino acid residues occurs in two forms without (Gastrin I) and

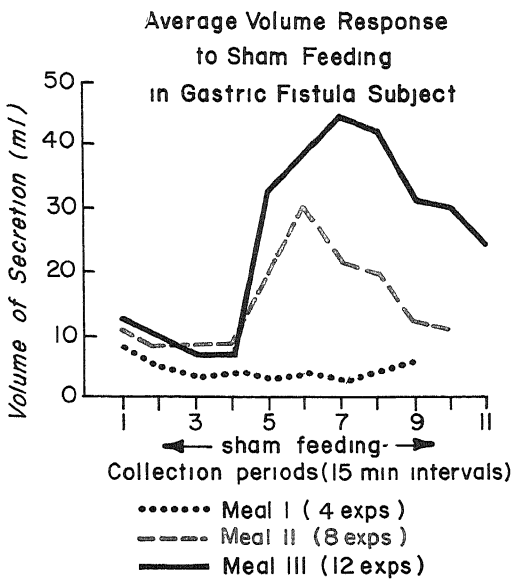


Fig 2 24 Volume response to sham feeding of different types of meals to a patient with a gastric fistula Meal I consisted of cereal gruel meal II was a routine hospital ward diet and meal III consisted of foods selected by the subject (Redrawn from Janowitz et al 1950)

with (Gastrin II) sulfate esterified on the phenolic hydroxyl group of the tyrosyl residue at position 12 Activity resides in the terminal tetrapeptide In the antrum and in the blood, gastrin exists also in a larger form than the 17 amino acid gastrin characterized by Gregory This later "big gastrin" appears to have the same physiological actions as the better known "little gastrin" Gastrin is present in small amounts, outside the antrum, and in the duodenum and probably in the pancreas

The action of gastrin on the acid-secreting cell is influenced by other gastrointestinal hormones and neural impulses Cholecystokinin and secretin inhibit gastrin stimulated acid secretion, the former probably by competitive inhibition, the latter by noncompetitive inhibition Cutting the vagus nerves greatly decrease the secretion of acid in response to gastrin

The release of gastrin from the pyloric gland area is markedly *inhibited* by acid bathing this area of the stomach A pH of 2 or less is required and can counteract the stimulatory effects both of vagal stimuli and of local distention and chemical factors

Interestingly enough, gastrin can be released by rapid increases in the concentration of serum calcium

These considerations help to elucidate the rationale for use of the operations of vagotomy, on the one hand, and of subtotal gastric resection on the other If the nervous and hormonal mechanisms of stimulation were independent of each other, we should expect that cutting the vagus nerves would influence only the initial reflex phase of gastric secretion and leave the gastric phase in full effect In reality, however, all phases of gastric secretion are profoundly depressed by vagotomy, and in some patients permanent achlorhydria has been reported Likewise, resection of the antrum, or subtotal gastric resection, theoretically should affect only the gastric phase Experience indicates, however, that the cephalic phase too is influenced profoundly

INTESTINAL PHASE

It has long been observed that the presence of certain food substances in the small intestine also stimulates gastric secretion This so-called intestinal phase has been extensively investigated The latent period of the intestinal phase is 2 to 3 hr but once secretion is initiated, it may last for 8 to 10 hr The following substances have been observed to excite gastric secretion when placed in the small intestine water, extractive substances of meat, products of protein digestion, (for example, peptone and amino acids), milk, alcohol, histamine, saponin, epinephrine, 0 1 N hydrochloric acid, 10% glycerin solution, and magnesium sulfate The intestinal phase is characterized by less secretory activity than either the cephalic or gastric phases and for this reason, the interpretation of results of studies of the intestinal phase have been more difficult

Experiments on the intestinal phase of gastric secretion have been conducted primarily on animals in the fasting state with the stomach empty Because of the long duration of the intestinal phase of gastric secretion from previous feedings, it is customary to fast the animal (dog) at least 24 hr The results obtained might have been otherwise had the intestinal phase been permitted to follow the cephalic and gastric phases as in the intact situation The known tendency of stimuli of different sorts to act synergistically in the gastrointestinal tract suggests, but does not prove, that the stimuli normally causing an intestinal phase may be much more effective in the presence of the re-

sidual stimuli of the cephalic and gastric phases

The cause of the intestinal phase is unknown, except that it is due to chemical stimuli A completely denervated gastric pouch secretes when food is placed in the intestine, proving that it is secreting in response to some chemical substance circulating in the blood The nature of the stimulus is unknown but may be a hormone, the "intestinal gastrin "

Perhaps one reason for the somewhat erratic behavior of the intestinal phase of gastric secretion is the ease with which inhibition of secretion may be induced by placing certain substances in the intestine The inhibitory effect of acid releasing secretion has already been mentioned, although acid in certain concentrations may stimulate gastric secretion, if the intestinal pH is lowered to 2 5 or below, it acts as an inhibitor

Intestinal (Duodenal) Inhibition of Gastric Secretion. It is appropriate at this point to consider some mechanisms inhibiting gastric secretion which originate in the upper intestinal tract. It has been well established that the presence of acid, fat, and hypertonic solutions in the duodenum released a hormone or hormones which inhibit gastric acid secretion To these hypothetical hormones, the term enterogastrone has been applied ("any hormone released from the intestine which inhibits gastric secretion," Johnson and Grossman, 1970)

While the evidence for an inhibitory action by ingested fats is conclusive, this enterogastrone has not been obtained in pure form It inhibits gastric motility as well The urinary substance(s) derived from this enterogastrone has been named urogastrone and much effort is being directed to its chemical elucidation

The inhibition of acid secretion which follows entry of HCl into the upper gastrointestinal tract is of great interest Part of it can be accounted for by the release of secretin (a known inhibitor of HCl secretion), which follows the presence of acid in the length of the entire duodenum, and appears to be related to the quantity of acid entering the gut Acid may inhibit by other mechanisms since Anderson, working in Uvnas's laboratory, has demonstrated that acidification of the duodenal bulb will inhibit gastrin and sham feeding stimulation of acid secretion without stimulating (secretin mediated) pancreatic secretion

INTERDIGESTIVE PHASE

Even after prolonged fasts, hydrochloric acid is found in the gastric juice of man as well as in that of the dog Both Pavlov and Babkin believed that hydrochloric acid was not secreted in the gastric juice during the interdigestive period When acid was detected in gastric juice of the dog after a prolonged fast, it was thought to result from either swallowing of saliva, regurgitation of duodenal juices, or extraneous stimuli such as the sight or smell of food In man, Heolzel (1926) found variable amounts of hydrochloric acid in gastric juice throughout a fast period of 40 days In both man and dog, the interdigestive secretion of HCl apparently occurs in an intermittent or periodic fashion The causes of such secretion are not known, but mechanisms involved apparently include both humoral and neural stimuli A subcutaneous autotransplanted gastric pouch in a dog, devoid of all original extrinsic nerve and vascular supply, during a fast will secrete acid intermittently This indicates that the stimuli for secretion must be humoral in nature The role of the neural pathways is suggested by the fact that following vagotomy in man, interdigestive secretion of hydrochloric acid is inhibited but not completely abolished With intact vagal innervation, conditioned reflexes in both man and the dog undoubtedly play a part In addition, in man, psychic stimulation is also probably important

The Normal Course of Gastric Secretion

With the discussion of the various phases of gastric secretion as background, we may now consider the sequence of events affecting the secretion of gastric juice following an "average" meal First, there is the sight and smell and the taste of food which, if agreeable, and if the individual has an appetite, will initiate the psychic or cephalic phase of gastric secretion This phase is augmented when the food is actually chewed and swallowed The first juice secreted has powerful digestive action, particularly upon the protein constituents of the food and releases from the food chemical excitants for the second, or chemical, phase of secretion These, along with the mechanical action of such food particles as enter the pyloric antrum, cause the release of gastrin. This augments and prolongs the secretion due to the cephalic phase

and initiates the gastric phase The gastric phase of secretion continues as long as food remains in the stomach As soon as the stomach is empty and the buffering substances present during digestion are no longer available, the pH of the antrum or the duodenum falls to a point at which the inhibitory effect of acid becomes evident, and that period of gastric secretion is brought to an end During the gastric phase of secretion, some chyme enters the small intestine where, after a variable latent period of 1 to 2 hr , it initiates the intestinal phase of gastric secretion This phase probably accounts for less than 20% of the total acid secreted by the stomach In duration, however, the intestinal phase exceeds the other phases, in fact, it may persist for as long as 5 hr Following the completion of the intestinal phase, the interdigestive phase occurs During the interdigestive phase, gastric secretion of minimal amount occurs intermittently

Effects of Some Drugs on Gastric Secretions

SECRETORY STIMULANTS

Histamine is one of the most powerful stimulants of gastric secretion Even endogenous histamine liberated within the body, e g , in dermographism, or even by the immersion of the hand in cold water at a temperature of 10° C, causes a detectable gastric secretory response within 15 min Administration of exogenous histamine is commonly used in gastric analysis Antihistamine agents, such as neoantergen and benadryl, do not inhibit gastric secretion evoked by histamine Histamine is effective when administered subcutaneously in very small doses Even a few 10ths of a milligram (0 2 to 0 3 mg of histamine base) will cause a copious secretion in dogs or man A similar dose given rapidly intravenously has little effect on gastric secretion If given very slowly, over a period of an hour, intravenous histamine is as effective as subcutaneous histamine in stimulating gastric secretion It is thought that histamine acts directly on the parietal cell to stimulate secretion of acid Large amounts of histamine are found in the gastric mucosa, and during active secretion, histamine can be detected in low concentrations in the gastric juice and the urine Histaminase, an enzyme that destroys histamine, is not found in the gastric mucosa

The effects of histamine on gastric secretion have been studied extensively There is a specific dose-response relationship, as illustrated by Code and associates (1949) in Figure 2 25 By increasing the dose of histamine in a step-like manner, the parietal cells can be stimulated to produce a "maximal secretory response " The latter is not equivalent to the *secretory capacity* of the stomach, for it has been shown that through the administration of a parasympathetic drug such as acetyl-β-methylcholine chloride (Mecholyl), and sham feeding, it is possible to stimulate secretion beyond the level achieved by prior histamine administration (Marks et al , 1960) Even so, the maximal secretory response to histamine has been shown to correlate well with the total number of parietal cells in the gastric mucosa Cox (1952) has attempted to calculate the number of parietal cells in the gastric mucosa on the basis of sampling representative fields using human autopsy material Accordingly, the average size of the parietal cell population is 1 18 billion in the adult male and 0 84 billion in the adult female In 15 patients with duodenal ulcer, the average number of parietal cells was 1 8 billion (1 9 billion for 11 males and 1 52 billion for 4 females) In 13 cases of gastric ulcer (sex not stated), the

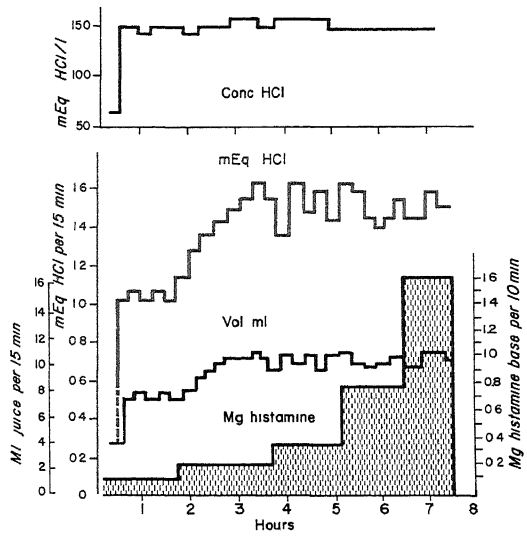


Fig 2 25 Dose-response relationship of a Heidenhain pouch to histamine given subcutaneously every 10 min Note that after a dose of 0 4 mg of histamine base/ 10 min was reached no further increase in gastric secretion occurred even with increasing amounts of histamine administration (Redrawn from Code et al 1949)

average number of parietal cells was 0 8 billion Card and Marks (1960) studied patients before and after partial gastrectomy for duodenal ulcer by measuring the maximal secretory response to histamine and counting the number of parietal cells in the portion of the stomach that was resected When the reduction in secretory response after partial gastrectomy was compared to the estimated number of parietal cells in the resected portion of the stomach, a good correlation was found and indicated that 1 billion parietal cells can secrete, on the average, 25 mEq HCl/hr (Fig 2 26) Marks and his associates (1960) have studied the maximal secretory response to histamine in the dog, and compared this to the "parietal cell mass " A linear relationship was demonstrated These investigators found that, in the dog, 1 billion parietal cells, when responding maximally to histamine, secreted on the average 9 8 mEq HCl/30 min

Histalog, an analog of histamine, (3-β-aminoethylpyrazol) is also a powerful gastric secretory stimulant It is frequently used in place of histamine for gastric analysis, as histalog does not produce the undesirable side effects caused by histamine

Pentagastrin, the commercially available terminal tetrapeptide of gastrin, is a powerful stimulant of gastric secretion and is being widely used as a clinical test substance A dose of 6 μg /kg intramuscularly or subcutaneously produces a peak acid output identical to those produced by a maximum dose of gastrin, histamine, or histalog

Caffein and *alcohol* are strong secretory stimulants The latter has a pronounced secretagogue action causing the secretion of a juice of high acidity and rich in mucin Alcohol exerts its secretory action through the liberation of gastrin

Parasympathomimetic agents as acetylcholine, Mecholyl (acetyl-β-methylchlorine chloride), pilocarpine, and nicotine are secretory stimulants

Insulin, in a dosage sufficient to lower the blood sugar to about 50 mg /100 ml , has an effect on the gastric glands similar to that of stimulation of the vagus nerves, namely, that the output of acid and pepsin is increased By an ill-understood mechanism, vagal stimulation also causes hypoglycemia Hypoglycemia activates sympathetic as well as parasympathetic centers Insulin convulsions are merely a further manifestation of this general excitatory

The Relationship Between the Maximum Acid Output and the Parietal Cell Mass

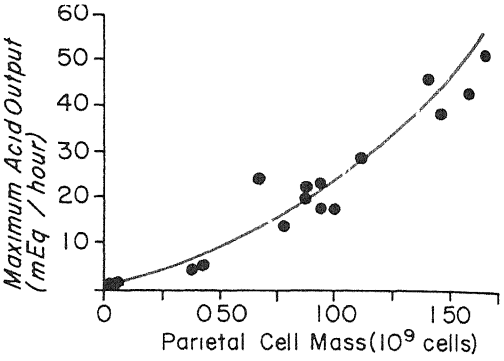


Fig 2 26 The relationship between the maximum acid output and the parietal cell mass calculated for the resected portion of each of 17 human stomachs The figure indicates that one billion parietal cells can produce approximately 25 mEq of HCl/hr (Redrawn from Card and Marks 1960)

effect of hypoglycemia of rapid onset Release of gastrin by the pyloric antrum may be reduced by insulin, as well as by vagal excitation, both of which are associated with hypoglycemia

Our understanding of the action of insulin on gastric secretion is complicated by observations which indicate that the hormone exerts an inhibitory effect in addition to the excitatory influence just described The duration of the inhibitory influence is brief (15 to 30 min) in man, but as long as 1 hr in dogs The inhibitory effect of insulin has been demonstrated during intravenous glucose administration, thus ruling out hypoglycemic mediation Insulin inhibition effectively controls gastric secretion induced by vigorous histamine stimulation

Since the effect of insulin on gastric secretion is dependent upon the integrity of the vagus nerves, the "insulin test" has been used as a means of determining the completeness of therapeutic vagotomy in human subjects (Hollander, 1946) A positive insulin test is proof of the presence of intact vagus fibers but a negative result is less conclusive, since some subjects with intact vagi fail to secrete in response to insulin The test, has, nevertheless, been of great value in the surgical control of duodenal ulcer.

SECRETORY DEPRESSANTS

Belladonna or its alkaloids, *atropine* and *hyoscine*, and *hyoscyamus* or its alkaloid *hyoscyamine*, are secretory depressants These

agents are capable of abolishing the cephalic phase of gastric secretion through their action on the vagus nerves Atropine reduces but does not suppress secretion due to histamine, alcohol, or caffeine Atropine has recently been shown not to inhibit the release of gastrin in response to feeding in man

Certain *synthetic anticholinergic agents* which inhibit gastric secretion with minimal side effects are available These agents are commonly used in the control of gastric acidity in patients with duodenal ulcer Many of these agents have a greater inhibitory action than atropine on gastric secretion and motility, and have less effect on other peripheral structures than atropine does, so that their use for regulation of gastric secretion becomes feasible In general, these parasympatholytic drugs can be divided into *tertiary* and *quaternary* amines The tertiary amines produce few side effects but usually are not too potent as secretory inhibitors The quaternary amines, on the other hand, are potent secretory inhibitors, and when given in adequate therapeutic doses, effectively suppress gastric secretion Such agents are particularly valuable in controlling the cephalic and interdigestive phases of gastric secretion Mild side effects as dryness of the mouth, blurring of vision, or slowing of the urinary stream are not uncommon, however

Effect of Emotional States on Gastric Secretion

In general, irrespective of the nature of the emotion, whether it be fear, anger, or frustration, the effect on gastric secretion is one of inhibition Beaumont (see 1929 republication) observed, in Alexis St Martin, that anger and fear caused suppression of gastric secretion In his book, *Bodily Changes in Pain, Hunger, Fear and Rage*, Walter B Cannon (1915) focused attention on the inhibitory effects of intense emotions That certain types of emotional disturbances stimulate gastric function, in fact increase secretion of hydrochloric acid, was first demonstrated by Wolf and Wolff (1943) in studies on their modern Alexis St Martin, Tom Little When exposed to sudden fear or pain, Tom's stomach became pale and ceased to secrete hydrochloric acid in accordance with the concept popularized by Cannon, however, when a situation developed which caused prolonged resentment or deep seated chronic anger, the opposite effects were observed, the gastric mucosa became congested and the secretion of hydrochloric acid was increased The weight of evidence indicates that chronic anxiety, anger, resentment, and fear, if sufficiently prolonged, tend to produce effects on gastric function characteristic of increased activity of the parasympathetic division of the autonomic system, including increased secretion of hydrochloric acid by the stomach

The central mechanism for these effects is situated partly in the hypothalamus This portion of the brain has been shown repeatedly to be a center for the coordination of impulses involved in emotional expression (Bard, 1928, 1929, Fulton and Ingraham, 1929)

It has long been known that stimulation of the posterior portion of the hypothalamus brings about visceral responses characteristic of the activity of the sympathetic division of the autonomic nervous system Evidence that the vagi, or other parasympathetic nerves, could be activated from the hypothalamus has not been so abundant, but it is becoming increasingly apparent that parasympathetic effects can be obtained from various regions of the hypothalamus by appropriate stimulating techniques (Strom and Uvnäs, 1950)

Measurement of Gastric Acidity

The words "free acid," "combined acid," and "total acid" were coined at a time when we had little understanding of the nature of gastric acidity, and little knowledge of ionization and the nature of buffers These terms were based on the concept that a certain fraction of the hydrochloric acid existed free in solution, a certain other fraction existed in combination with protein or other acid-absorbing substances, and that these fractions could be distinguished by the use of indicators We realize now that indicators merely tell us when the solution has reached a certain pH It is customary now to measure titratable acidity by titration with 0 1 N NaOH to a fixed end point (pH) colorimetrically or electrometrically, and to measure hydrogen ion activity with a pH meter

"Titratable acidity = hydrogen ion concentration + un-ionized hydrogen concentration" is the current equation instead of the obsolete terms "Total acidity = free acidity + combined acidity "

Much discussion has been devoted to the question of what is the correct end point for titratable acidity Some have suggested pH 7 0

as physiochemical neutrality; others recommend pH of plasma of 7.4.

It is customary to measure the pH of the gastric contents using a glass electrode and pH meter. pH, it will be remembered, is the negative logarithm of the hydrogen ion concentration. The glass electrode measures the hydrogen ion activity, not true concentration, and can be converted to concentration by a correction factor, the activity coefficient. Thus, from the pH value, one can calculate the hydrogen ion concentration in terms of moles per liter. The pH value is particularly useful when the acidity is low (James 1957).

It should be remembered that expression of gastric acidity in terms of free acid and total acid is an expression of the *concentration* of HCl in the gastric contents. Quantitative data of output can be expressed in terms of *milliequivalents* of hydrochloric acid secreted *per unit of time*.

6 / PANCREATIC SECRETION

Structure of the Pancreas

The pancreas is both an endocrine and exocrine gland It develops from two outgrowths (ventral and dorsal) of the primitive gut The ventral pancreatic bud arises with the hepatic diverticulum, and together they grow into the ventral mesentery The dorsal pancreatic bud enters the dorsal mesentery, and, as a result of the rotation of the gut together with its mesenteries as well as growth of the pancreatic anlages themselves, the two portions of the pancreas meet along the left border of the duodenum and soon fuse with one another The duct of the ventral pancreas becomes confluent with that of the dorsal anlage, usually near its origin In keeping with the fact that the proximal portion of this duct was developed in conjunction with the liver bud, the definitive pancreatic duct (of Wirsung) enters the duodenum in close association with the common bile duct That portion of the duct of the dorsal rudiment which lies between the duodenum and the point of anastomosis with the ventral duct is normally retarded in subsequent development and becomes the accessory duct or *duct of Santorini*

The pancreas is a racemose gland, its alveoli resembling those of the salivary gland in their general arrangement and design The pancreas is made up chiefly of groups of cells forming acini which tend to be spherical or ovoid in general, in the closely packed tissue of the pancreas they are actually polygonal Groups of acini form primary lobules, also polygonal in contour, which are imperfectly separated from other primary lobules by incomplete connective tissue septa Numerous adjacent primary lobules form a secondary lobule The secondary lobules are completely surrounded by connective tissue The pancreatic tissue proper is composed of acinar cells, islet cells, and duct cells The acinar cells tend to be pyramidal, with the truncated apex of the pyramid directed toward the lumen They are large with well-developed nuclei and abundant granular cytoplasm The granules (zymogen granules) vary in number and position in the cells depending upon the activity of the gland, but tend to be more abundant in the apical region of the cell which is adjacent to the lumen of the acinus

Cells, corresponding in structure to the cells of the terminal ducts, may be seen in microscopic sections within the acini, these are called *centroacinar cells* and contain abundant mitochondria

INNERVATION OF THE PANCREAS

The pancreas receives its nerve supply from the vagi and splanchnic nerves The extrinsic nerve supply to the pancreas passes through the celiac plexus and reaches the gland by way of the nerve plexus surrounding the arteries, chiefly the superior pancreaticoduodenal artery The vagus fibers end in the intrinsic ganglia of the pancreas From these the parasympathetic path is continued through postganglionic unmyelinated fibers to the acinar cells, islet cells, and the smooth muscles of the ducts The efferent splanchnic fibers, destined to supply the pancreas, all end in the celiac and associated ganglia From there the sympathetic path is continued through unmyelinated, postganglionic fibers which are distributed solely to the pancreatic blood vessels Myelinated visceral afferent fibers from the pancreas are also present in the splanchnic nerves

It is often incorrectly stated that the pancreas receives all, or the greater part, of its vagus nerve supply from the right vagus This error arises from the tendency to overlook the fact that after the vagi merge in the esophageal plexus, the right and left nerves can no longer be distinguished Two or more trunks emerge from the esophageal plexus The trunk that occupies the more posterior position at the level of the diaphragm, although referred to by some anatomists as the right vagus, contains fibers from both the right and left vagi as they occur in the neck When stimulated in the neck, the right and left vagi are equally effective in modifying pancreatic function

Pancreatic Juice

The pancreas secretes a colorless, odorless, alkaline fluid of low viscosity Exceptional specimens having a high concentration of enzymes may be more viscous Pure pancreatic juice collected from dogs without loss of carbon dioxide has a pH between 8 0 and 8 3, and a specific gravity that ranges from 1 007 to 1 042, but usually falls between 1 010 and 1 018

Pancreatic juice is iso-osmotic with the plasma, and its osmolality is independent of rate of flow Induced alterations in the osmotic pressure of the plasma are followed rapidly by corresponding alterations in the juice

ELECTROLYTE COMPOSITION

Sodium concentration in the juice is approximately equal to Na^+ in plasma water, as is K^+ concentration both are independent of the rate of flow (Fig 2 27) Table 2 8 (Janowitz, 1967) records the values for dog pancreatic juice The values for man are similar Calcium concentration is lower (1 7 ± 0 3 mmol /kg H_2O) than in plasma Calcium appears to enter pancreatic juice along with the enzymic components, not with the major electrolyte (HCO_3^-) containing fluid

It is now well established that the bicarbonate and chloride concentrations of pancreatic juice vary in a reciprocal manner, so that the sum of the two expressed in milliequivalents is approximately constant and nearly the same as the total cation of the blood plasma In human pancreatic juice, Janowitz and Dreiling (1962) have found the sum of bicarbonate and chloride concentrations to remain remarkably constant at different secretory rates They reported a value of 154 ± 10 mmoles/kg H_2O

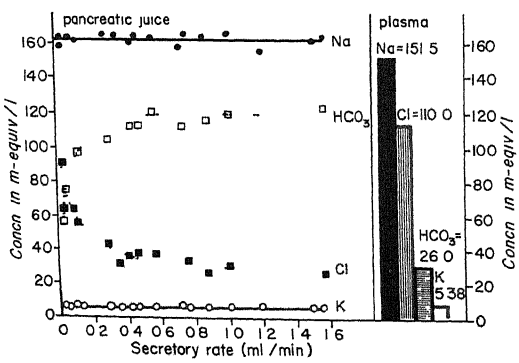


Fig 2 27 Relation between rate of secretion and concentrations of electrolytes (From Bro-Rasmussen et al 1956)

Table 2 8
Electrolytes of dog serum and pancreatic juice

	Serum mmole/kg H_2O	Juice mmole/kg H_2O
Cl	117 8 ± 4 8	
Total CO_2	28 3	154 ± 10
Na^+	153 1 ± 3 2	154 ± 7
K^+	5 1 ± 0 3	4 8 ± 0 9
H_2O	92 4%	98%

From Janowitz (1967) p 926

The bicarbonate concentration increases and the chloride concentration decreases with increasing rates of secretion This relationship holds over a wide range of secretory rates and is such that a straight line results when the logarithm of the bicarbonate concentration in milliequivalents/liter is plotted against the *reciprocal* of the rate of secretion expressed as cubic centimeters of secretion per kilogram of body weight per minute, the slope indicates an inverse relation The maximum bicarbonate concentration in juice collected from different dogs ranges between 135 and 148 mEq /liter and the maximum rate of secretion is about 0 1 ml /kg /min , or about 1 ml /min for a 10-kg dog, higher rates than this are occasionally seen The relationship of bicarbonate concentration to flow rate in pancreatic juice collected from man is illustrated in Figure 2 28 In addition to chloride and bicarbonate, the pancreatic juice contains a small amount of phosphate, but less than that occurring in blood plasma The electrolyte composition of a sample of human pancreatic juice secreted at a relatively high rate is illustrated in Figure 2 29

PROTEIN CONTENT

The protein content of pancreatic juice varies over a wide range, but in an orderly manner with the conditions governing secretion In samples of dog's pancreatic juice collected under different conditions, the protein content varied between 0 1 and 10% Concentrations found in human pancreatic juice collected through fistulas are nearer the lower extreme, ranging from 0 1 to 0 3% The protein in the juice consists mainly of enzymes Using electrophoresis, it has been possible to determine that there are at least 10 distinct proteins in pancreatic juice, with the possibility of still others that fail to separate out in the time during which the electrical field is maintained (Gross-

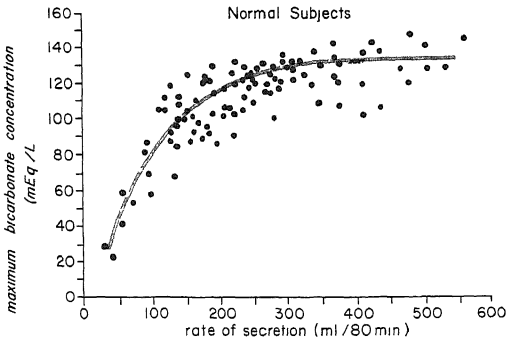


Fig 2 28 Maximum bicarbonate concentration observed at different rates of pancreatic secretion in normal subjects following stimulation of the pancreas with intravenous secretin (Redrawn from Dreiling and Janowitz 1959)

berg, Komarov, and Shay, 1952) This method of separating the proteins of pancreatic juice also separates the enzymatic activity, proteolytic activity is found in one component, lipolytic activity in another, and amylolytic activity in another In man, the following enzyme proteins have been reported amylase, lipase, trypsinogen, chymotrypsinogen, procarboxypeptidase A and B, ribonuclease, deoxyribonuclease, proelastase, and trypsin inhibitor

PANCREATIC ENZYMES

The enzymes of pancreatic juice are capable of digesting all three classes of foodstuff The proteolytic enzymes, as they occur in the pancreatic tissue and freshly secreted pancreatic juice, possess no proteolytic activity and are referred to in that state as *proenzymes* and require activation For trypsin the proenzyme is trypsinogen Kunitz and Northrup (1934) have shown that what was called trypsin or trypsinogen consists of at least two components, for one of which they retain the name *trypsin* and for the other they coined the name *chymotrypsin* The proenzyme of chymotrypsin is known as chymotrypsinogen

Trypsinogen. The molecular weight of *trypsinogen* is approximately 25,000 Trypsinogen is a single polypeptide chain containing 229 amino acid residues (Neurath, 1964) Cleavage of trypsinogen between residues 6 and 7 yields active trypsin. As an enzyme trypsin hydrolyzes native protein, producing proteoses and peptides it accelerates coagulation of the blood, in which it acts as a thrombokinase In its action on proteins, trypsin hydrolyzes only those peptide bonds whose carbonyl group is contributed by an amino acid that has a positively charged side group as arginine or lysins Trypsin cleaves peptide bonds that occupy an internal position in a polypeptide chain and also is referred to as an "endopeptidase" Its optimum pH for casein lies between pH 8 and 9

Trypsinogen is indistinguishable on chemical analysis from trypsin, but has different solubilities and a different crystalline form, and it exhibits no proteolytic activity It changes spontaneously in solution to active trypsin, but the change is accelerated by acid or by exposure to concentrated magnesium sulfate, calcium chloride, active trypsin, or to certain activating enzymes called *kinases* The effectiveness of trypsin in the activation of trypsinogen is of great importance because such autocatalytic reactions progress with great rapidity once they are started and require only minute amounts of activating agents to set them going Spontaneous activation of trypsinogen is suppressed in crude extracts (and doubtless also in living pancreas) by the presence of a trypsin inhibitor Trypsinogen is activated on contact with the intestinal mucosa by an enzyme in the mucosa called *enterokinase* Enterokinase does not appear in the intestinal secretion in the absence

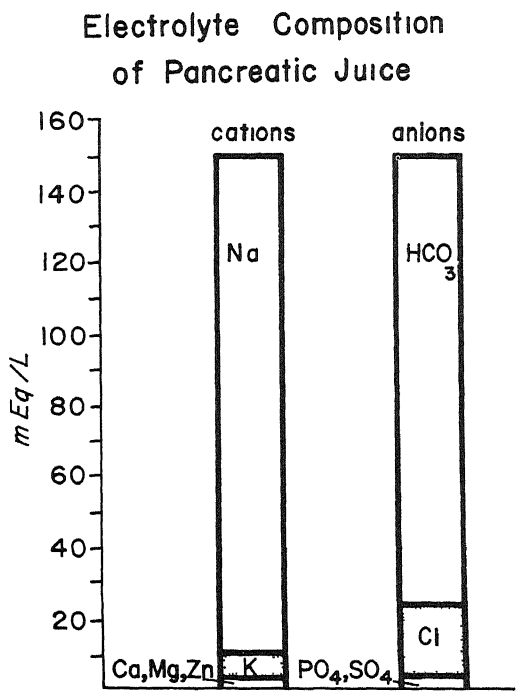


Fig 2 29 Example of the electrolyte composition of human pancreatic juice secreted at a high flow rate of 3 5 ml/min

of pancreatic juice but is present in intestinal mucosal extracts

Chymotrypsinogen Chymotrypsinogen has a molecular weight of approximately 25,700 and, like trypsinogen, is a single polypeptide chain containing 246 amino acid residues, or 17 more than trypsinogen (Neurath, 1964) When cleaved between residues 15 and 16, chymotrypsinogen is converted to the active enzyme chymotrypsin The "active centers" of both trypsin and chymotrypsin are remarkably similar In each enzyme a "histadine region" and a "serine region" have been localized and found to be connected by disulfide bonds (from two half-cystine residues) that lock the histadine and serine residues in a fixed position

Chymotrypsinogen is converted to chymotrypsin by active trypsin only, but indirect activation can be accomplished by enterokinase if trypsinogen is also present If chymotrypsinogen is activated rapidly by fairly large amounts of trypsin, π-chymotrypsin is formed which then changes spontaneously to δ-chymotrypsin, as a result of the π-chymotrypsin acting upon itself If smaller amounts of trypsin are used the major products are α-, β-, and γ-chymotrypsin which appear to have essentially the same molecular weight as chymotrypsinogen and π- and δ-chymotrypsin A sixth type, known as chymotrypsin (and chymotrypsinogen) β, also has been isolated The activity of chymotrypsin is comparable to that of trypsin but it has different specificities, for example, it coagulates milk but not blood Chymotrypsin digests casein somewhat more rapidly than does trypsin, but the digestion is carried much further by a mixture of the two enzymes than by either alone, indicating that the two enzymes attack different linkages in the protein molecule Chymotrypsin acts preferentially on peptide bonds whose carbonyl group is adjacent to an amino acid residue that has a six-carbon ring in its side group, as tyrosine and phenylalanine Like trypsin, chymotrypsin is also an "endopeptidase," and cleaves peptide bonds that occupy an internal position in a polypeptide Chymotrypsin is reversibly inactivated by boiling in acid solution Its optimum activity occurs in the same pH range as that of trypsin

Other Proteolytic Enzymes. The peptides resulting from protein digestion by trypsin and chymotrypsin are further digested by enzymes known as *peptidases* Most of these come from the intestinal mucosa, but two have been identified and isolated from pancreatic juice enzymes known as *carboxypeptidase* A and B, "exopeptidases" which split off the end amino acid containing a free carboxyl group from polypeptides In fresh pancrease, they occur in an inactive form, procarboxypeptidase A and B, and, like chymotrypsinogen, are activated by trypsin

Pancreatic juice also contains *ribonuclease* and *deoxyribonuclease*, which partially hydrolyze the corresponding nucleic acids into mononucleotides The optimum pH for deoxyribonuclease is in the neighborhood of 7 *Elastase* (Balo and Banga, 1949), and more recently, *collagenase* (Houck and Patel, 1962), have also been demonstrated in pancreatic juice

Lipase *Pancreatic lipase* accounts for the ability of the juice to emulsify and hydrolyze fats This enzyme is known principally through its activity and has not been isolated in a pure state or crystallized Pancreatic lipase hydrolyzes fat only in the presence of bile salts or other substances with similar properties, for example, some of the synthetic detergents Bile salts in high concentration inhibit the action of lipase, as do salts of certain heavy metals, e g , copper, iron, and cobalt, it is also inhibited by the halogens, fluorine, iodine, and bromine The pH at which pancreatic lipase exhibits its optimum activity varies with the substrate, but is always on the alkaline side of neutrality Pancreatic lipase hydrolyzes triglycerides to free fatty acids and glycerol

Pancreatic juice also contains two phospholipases *Phospholipase A* splits a fatty acid off lecithin or cephalin to form lysolecithin or lysocephalin *Phospholipase B* is capable of splitting a fatty acid off lysolecithin to form glyceryl-phosphoryl-choline

Amylase. *Pancreatic amylase* is an α-amylase, as is salivary amylase Amylase has been isolated in crystalline form from human pancreas The molecular weight is approximately 45,000 Pancreatic amylase splits an α-1,4-glycosidic bond to hydrolyze starch to maltose The amylolytic activity has an electrophoretic mobility in human serum similar to that of γ globulin Amlyase is excreted in the urine and the amylolytic activity of urine is an excellent indicator of the amount of amylase released from the pancreas into the blood Pancreatic amylase is secreted in active form but becomes inactive if dialyzed against distilled water—

owing to the removal of neutral salts which are thus proved to be essential for the activity of the enzyme The most efficient activating ion is chloride The optimum pH for activity of pancreatic amylase varies between 6 9 and 7 1 The enzyme digests glycogen as well as starch, and under favorable conditions, the end product in both instances is maltose

The importance of the pancreatic enzymes in digestion is evidenced by the fact that when pancreatic juice is excluded from the intestine by pancreatectomy or ligation of the ducts, there is serious impairment of digestion and absorption of the various foodstuffs This is most pronounced in the case of fat and protein, unabsorbed fat particularly appearing in large amounts in the stools The bicarbonate of the juice also serves the neutralize the gastric juice as it enters the intestine and to establish in the intestine a pH more favorable for intestinal digestion than would otherwise be the case

Mechanism of Pancreatic Secretion

It has long been assumed that the inorganic constituents of pancreatic juice are secreted by the same cells that produce the enzymes, namely the acinar cells However, the fact that fluid secretion and enzyme secretion are separately regulated raises a theoretical objection to this view All current theories of stimulation require that cells shall respond to a stimulus with a characteristic type of functional activity regardless of the nature of the stimulus In the other digestive glands, about which we have more information, each component of the secretion that is independently regulated is formed by a different type of cell We must also take account of the clinical observation that in certain types of pancreatic disease, enzyme secretion may be depressed or absent, whereas secretion of fluid and bicarbonate remains normal (Lagerlöff, 1939, Friedman and Snape, 1950) On the other hand, in glands injured by alloxan, fluid secretion may be subnormal, whereas enzymes are secreted normally This latter observation, reported by Grossman and Ivy (1946), contains a hint as to a possible solution of the problem Histological examination of the gland which failed to respond normally to secretin showed no injury to the acinar cells, but the cells of the intralobular ducts appeared abnormal, in particular, they exhibited vacuolization The intralobular duct cells are continuous with the centroacinar cells that line the alveoli and are identical with these cells in their cytological structure It is now thought that cells of this type secrete water and bicarbonate and that the acinar cells secrete the enzymes

Finally, the enzyme carbonic anhydrase has been demonstrated by histochemical methods only in ductular epithelium, and in view of the role of carbonic anhydrase in the secretion of HCO_3, suggests further that this structure is involved in the elaboration of water and electrolytes

In discussing details of the mechanism of pancreatic secretion, the secretion of fluid and electrolytes will be considered separately from the secretion of enzymes The evidence presented above indicates that pancreatic juice like gastric juice, is a mixture of secretory products from different glandular structures that are under the influence of separate regulatory mechanisms

FLUID AND ELECTROLYTE SECRETION

Three hypotheses have been suggested to account for the variation in composition of the electrolytes of pancreatic juice (See Fig 2 30)

1) Admixture concept the pancreatic juice is visualized as a mixture of three fluids neutral chloride at 174 mEq /liter, HCO_3 at 174 mEq /liter, and a third fluid containing enzymes, mucus, and an acid chloride

2) Exchange mechanism the pancreatic juice is formed by the elaboration by the centroa-

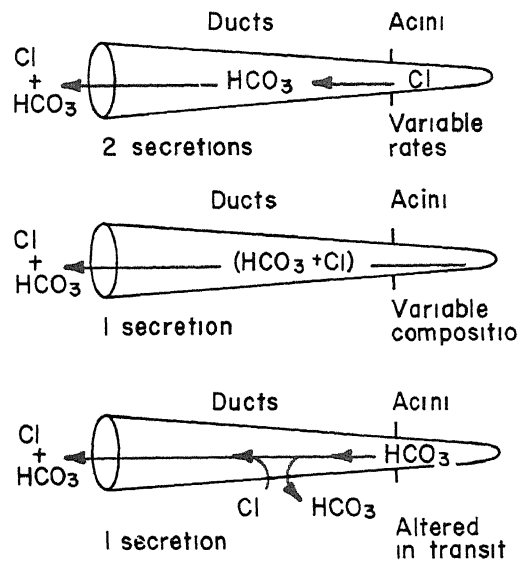


Fig 2 30 Hypothetical mechanisms of pancreatic secretion (From Makhlouf and Blum 1970)

cinar cells of a HCO_3 solution (isotonic with the plasma, 150 to 160 mEq /liter) and exchange of HCO_3 for Cl as the fluid traverses the pancreatic collecting ducts

3) A unicellular concept in which a single source elaborates a juice which the concentration of HCO_3 and Cl vary

The available data do not permit any conclusive choice among these alternatives, although perfusion and stop flow studies indicate that HCO_3 and Cl can exchange passively Micropuncture experiments are difficult to reconcile but some do indicate that HCO_3 is added to the juice along the smallest radicles of the ductular system

The bicarbonate in pancreatic juice is thought to come from both metabolic CO_2 and HCO_3^- from the blood plasma The pancreas contains an abundance of carbonic anhydrase, suggesting that it plays a role in the formation of bicarbonate It has been demonstrated that acetazolamide (Diamox), a potent carbonic anhydrase inhibitor, is capable of inhibiting the secretion of bicarbonate by the pancreas (Dreiling and Janowitz, 1959)

If the concentration of sodium or potassium is increased in the plasma by injecting hypertonic NaCl or KCl, the concentration of these cations is increased in the pancreatic juice to the same degree (Ball, 1930) In addition, if radioactive sodium is administered intravenously, it appears in pancreatic juice within 3 min, and after 15 min the activities of the isotope in pancreatic juice and serum are essentially the same (Montgomery, Sheline and Chaikoff, 1941)

Discussion so far has stressed the role of transport of the anions (HCO_3 and Cl), however, some recent evidence suggests that an active role for sodium transport may also exist The enzyme, adenosine triphosphatase (ATPase), shown to be involved in the active transport of Na in the kidney and elsewhere has been demonstrated in the pancreas, and inhibitors of this (Na, K) ATPase, such as ouabain, inhibit pancreatic secretion of fluid significantly It may be that the secretion of Na and HCO_3 are coupled

SECRETION OF ENZYMES

Secretion of enzymes by the pancreas involves at least three distinct processes *1*) synthesis within the cell of the enzymes or their synthesis within the cell of the enzymes or their precursors, *2*) storage of the enzymes within the cell in the form of zymogen granules, and *3*) discharge of the enzymes from the secretory cells into the ducts

Synthesis is a continuous process, dependent upon the availability of various amino acids that are required for the synthesis of the enzymes The nature of the dietary protein, and in particular its amino acid composition, has a profound effect on the total concentration of enzymes in the pancreas

Interesting light has been shed on the synthesis of enzymes by the studies of Hokin (1951, 1952) and Hokin and Hokin (1956) on pigeon and mouse pancreas in vitro They found that pigeon pancreas slices would synthesize amylase in vitro in glucose solutions, in serum, in amino acid solutions, and in saline Synthesis was most rapid in solutions of amino acids and least rapid in normal saline, it proceeded at an intermediate rate in serum One can see here the effect of an increased concentration of reacting substances (amino acids) on synthesis of protein by the cells of the pancreas Although pancreozymin and parasympathomimetic drugs caused discharge of amylase from the cells into the surrounding medium (secretion), the rate of synthesis of the enzymes was not materially affected Certain amino acids were essential for synthesis but others could be supplied from sources in the medium Those essential for maximum synthesis were tryptophane, arginine, threonine, valine, tyrosine, lysine, leucine, histadine, isoleucine, and phenylalanine

The synthesized material appears as *zymogen granules*, which accumulate in the resting gland until they come to occupy most of the available space within the cells (Hirsch, 1939, Hirsch et al , 1957, Sluiter, 1944) Under the influence of appropriate stimuli the zymogen granules diminish in number and in size, at the same time enzymes appear in the secretion Stimuli such as HCl and secretin, which produce a watery secretion with a low enzyme content, have little effect on the granules, whereas stimuli which call forth secretion rich in enzymes deplete the granules and, if they act over a long period of time, may leave the cells almost free of zymogen granules

According to the observations of Hokin and Hokin (1956) on pancreatic and salivary gland slices in vitro, effective stimuli for secretion of enzymes cause a remarkable increase in the

rate of turnover of phosphorus and its incorporation into the other soluble phospholipids, possibly in the phospholipids of the plasma membrane

Palade, Siekevitz and Caro (1962) have used electron microscopy and autoradiography to study intracellular synthesis, storage, and transport of enzymes Radioactive carbon-labeled leucine was injected into guinea pigs Pancreatic tissue was obtained at different time intervals for examination by electron microscopy and autoradiography Within 3 min , high specific activity was found in the ribonucleic acid (RNA) particles or ribosomes attached to the α-cytomembranes (endoplasmic reticulum) From the ribosomes, α-chymotrypsinogen of high specific activity was obtained by means of column chromotography The labeled enzyme is then transferred into cisternae formed from the α-cytomembranes These structures have been termed "rough-surfaced cisternae," due to the attached ribosomes Just how the enzyme is transferred from the ribosome across the cisternae membrane is not known In the region of the Golgi apparatus, the enzyme content of the rough-surfaced cisternae is transferred to smooth-surfaced vacuoles This occurs about 20 min after the administration of the labeled amino acid It is in the smooth-surfaced vacuoles that the enzyme is concentrated and stored The filled vacuoles are recognized by light microscopy as "zymogen granules" The stage of storage in a "mature" zymogen granule is reached in about 45 min In this form, enzymes can remain in storage for hours until discharged into the acinar lumen The process of discharge of the contents of the zymogen granules is thought to involve a coalescence of the membrane of the zymogen granule with the plasma membrane of the cell The various stages of the process just described are illustrated in Figure 2 31 The pancreatic acinar cell is capable of secreting enzymes at a high rate, even after zymogen granules have been depleted Thus other pathways than via storage in a zymogen granule probably exist for the secretion of pancreatic enzymes

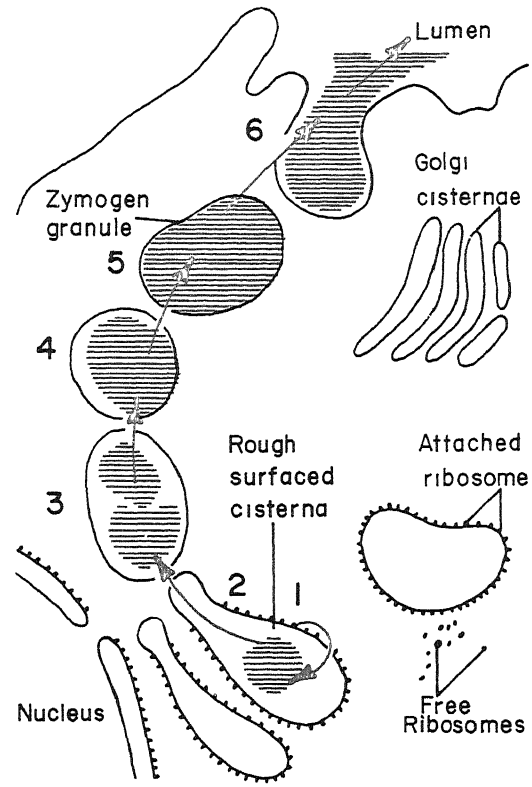


Fig 2 31 Diagram illustrating the various stages of synthesis storage and secretion of pancreatic enzymes by the acinar cell of the guinea pig *1* enzymes are synthesized by the ribosomes and transported into a rough-surfaced cisterna *2* enzymes are transferred from the rough surfaced cisterna into a smooth surfaced vesicle *3* and *4* near the Golgi apparatus enzymes are concentrated within the smooth-surfaced vesicles to become *5* a mature zymogen granule *6* illustrating the secretion' of enzymes into the lumen of the acinus (Redrawn from Palade et al 1962)

Control of Pancreatic Secretion

The pancreas is under the control of both *nervous* and *humoral* mechanisms The nervous control of pancreatic secretion is provided by both the sympathetic and parasympathetic divisions of the autonomic nervous system In addition, there is evidence that a *local cholinergic mechanism*, independent of vagal innervation, may play a role in controlling pancreatic secretion Humoral control of pancreatic secretion is provided by the hormones *secretin*, *pancreozymin*, and *gastrin*

NEUROGENIC CONTROL

The effects of *sympathetic stimulation* on pancreatic secretion are variable Both augmentation and decrease in flow rate have been reported Atropine is said to block the effects of sympathetic stimulation, causing an increase in flow rate, indicating that cholinergic secretory fibers are present in the sympathetic innerva-

tion of the pancreas As in the stomach, sympathetic stimulation can influence pancreatic secretion indirectly, due to changes in blood flow through the organ Sympathetic stimulation may also alter the flow rate due to effects on the pancreatic ducts, since constriction of the ducts would slow the passage of pancreatic juice

Stimulation of the vagus or *parasympathetic* innervation of the pancreas results in the secretion of enzymes but has little or no effect on the secretion of bicarbonate Acetylcholine is the mediator of the vagal effect Vagotomy and atropine markedly depress the secretion of enzymes by the pancreas The effects of vagal stimulation on pancreatic secretion can be demonstrated easily in the experimental animal In a dog with an esophagogastrostomy and pancreatic fistula, sham feeding produces a prompt secretion of enzymes by the pancreas Insulin hypoglycemia, known to produce vagal stimulation, also results in the secretion of enzymes Electrical stimulation of the peripheral segment of the cut vagus produces a similar response Cholinergic drugs, as pilocarpine, stimulate pancreatic secretion of enzymes whereas atropine and anticholinergic agents produce inhibition However, these kinds of stimuli may release gastrin from the antrum or upper gastrointestinal tract as well

The presence of a gastropancreatic reflex has been demonstrated by White, Lundh and Magee (1960) In dogs, these investigators found that distending the fundus of the stomach with a balloon caused an increase in volume and enzyme output in pancreatic juice Vagotomy abolishes the gastropancreatic reflex for a few weeks but then it returns, suggesting that local reflex pathways exist which are independent of vagal innervation

HUMORAL CONTROL

The two hormones, *secretin* and *pancreozymin*, provide for the humoral control of pancreatic secretion Both of these hormones can be extracted from the mucosa of the duodenum and upper small bowel The specific cells in the mucosa that produce secretin and pancreozymin are not known Under normal circumstances, secretin and pancreozymin are released into the blood by appropriate stimuli and carried to the pancreas to produce their effects Dilute hydrochloric acid in the duodenum is a potent stimulus for the release of secretin Peptones or products of protein digestion serve equally well for the release of pancreozymin Secretin acts on the ductule or centroacinar cells to stimulate them to secrete water and bicarbonate while pancreozymin stimulates the acinar cells to secrete enzymes

Secretin The discovery of secretin in 1902 by Bayliss and Starling was an event of great historical significance, inasmuch as secretin was the first chemical substance known to be secreted into the blood and act as a stimulus to an organ at a distance from the site of its formation Recognizing this fact, Bayliss and Starling coined the word *hormone* to characterize this and similar substances Before the discovery of secretin, it was known that acid in the intestine induced secretion of pancreatic juice in animals in which all the nerves to the pancreas and to the intestine had severed While workers in Pavlov's laboratory in St Petersburg were busily searching for the unknown nervous pathways which, they supposed, conveyed impulses to the pancreas after all the known fibers had been cut, Bayliss and Starling in England did the simple experiment of making an acid extract of the mucous membrane of the intestine and injecting it into an animal This caused the pancreas of the injected animal to secrete The excitement in Starling's laboratory can well be imagined A "chemical reflex" had been discovered, something hitherto unknown in physiology They called the new substance *secretin*

In the introductory section the structural formula of the polypeptide secretin with its 27 amino acids has been presented In contrast to gastrin, the entire molecule is required to stimulate the pancreas to secrete a bicarbonate secretion at high rates of flow

Pancreozymin. The crude extract prepared by Bayliss and Starling caused the pancreas to secrete enzymes as well as water and bicarbonates Since purified secretin seemed to stimulate only the output of water and bicarbonate, a search was made by Harper and Raper (1943) for the material in the extracts which stimulated secretion of the enzymes They succeeded in separating a fraction which did not increase the volume of pancreatic juice but did greatly increase the enzyme concentration This substance they called *pancreozymin*

The formula of this 33 amino acid polypeptide has already been presented, which has the same terminal tetrapeptide as gastrin, is iden-

tical with cholecystokinin, and causes the pancreas to release its enzymic content with little fluid

It has long been known that hydrochloric acid in the intestine liberates secretin into the blood, thus causing the pancreas to secrete The flow of pancreatic juice is also increased by distilled water, hydrochloric acid, soaps, and peptones The enzyme concentration is also increased by these same stimuli, thus proving that both secretin and pancreozymin are released by all the ordinary stimuli used to excite pancreatic secretion

INTERACTION OF NERVOUS AND HUMORAL CONTROL MECHANISMS

Although the various nervous and humoral control mechanisms have been discussed as separate or individual factors capable of influencing pancreatic secretion, the student should not be misled in thinking each mechanism acts independently of all others The pancreatic response to a meal indicates that in all likelihood there is a high degree of interdependence between the neural and humoral mechanisms

In a recent review, Preshaw (1967) concluded "mainly from studies with anticholinergic drugs" that cholinergic mechanisms contribute to the action of secretin on pancreatic secretion, but it is uncertain whether this mechanism is independent of the vagus He found little evidence to indicate that the release of pancreozymin involved similar cholinergic mechanisms Recognition that gastrin stimulates pancreatic enzyme secretion reinstates the vagus in this chain Potentiation between neural and hormonal mechanisms of pancreatic secretion have not been established

Tests of Pancreatic Exocrine Function

Pancreatic exocrine insufficiency may be caused by a number of diseases affecting the pancreas Since the pancreas is the principal source of lipase furnished to the upper small bowel for the digestion of fats, an insufficient amount of pancreatic juice and its contained lipase results in maldigestion of fats and causes steatorrhea The gastric juice provides pepsin and secretions from the small bowel (succus entericus) provide proteolytic enzymes so that maldigestion of proteins is not as marked as the defect in fat digestion Amylase provided by the salivary glands and various disaccharidases provided by the small intestines also promote digestion of carbohydrates, even in severe exocrine insufficiency of the pancreas Thus, the principal clinical manifestation of pancreatic exocrine insufficiency is an excess of fat excreted in the feces

FECAL FAT EXCRETION

A commonly used test for the detection of pancreatic exocrine insufficiency is the quantitative measurement of fecal fat excretion Normal individuals on an average diet seldom excrete more than 6 or 7 g of fat in the feces in a 24-hr period Patients with exocrine insufficiency of the pancreas may excrete up to 50 g of fat/24 hr , depending upon the dietary fat intake and the severity of the insufficiency

SECRETIN AND PANCREOZYMIN TESTS OF PANCREATIC FUNCTION

Since secretin and pancreozymin in injectable form have become available, tests of pancreatic function using these hormones as stimuli have been developed Secretin was first available in a form pure enough for use in humans and more experience has accumulated with it The secretin test was introduced by Ågren, Lagerlöff, and Berglund (1936) and has been used extensively in this country and abroad It is wise to test the patients for hypersensitivity to the secretin preparation before conducting the test To perform the test, a double-lumen tube is passed through the patient's nose or mouth, into the stomach for draining gastric juice, and the other into the duodenum for collecting the pancreatic and duodenal secretions It is important that the double-lumen tube be positioned accurately by means of fluoroscopy The gastric and duodenal secretions are collected separately Gastric juice is usually aspirated continuously by a low pressure vacuum pump while the duodenal contents are aspirated periodically Samples are collected during 10 or 20 min intervals After a control period of approximately 30 min , to provide for clearing of the duodenal samples of gastric juice, secretin, in a dose of 1 clinical unit/kg body weight, is given slowly intravenously over a 4- to 5-min period The duodenal contents are then collected in separate 10- to 20-min samples for a period of time,

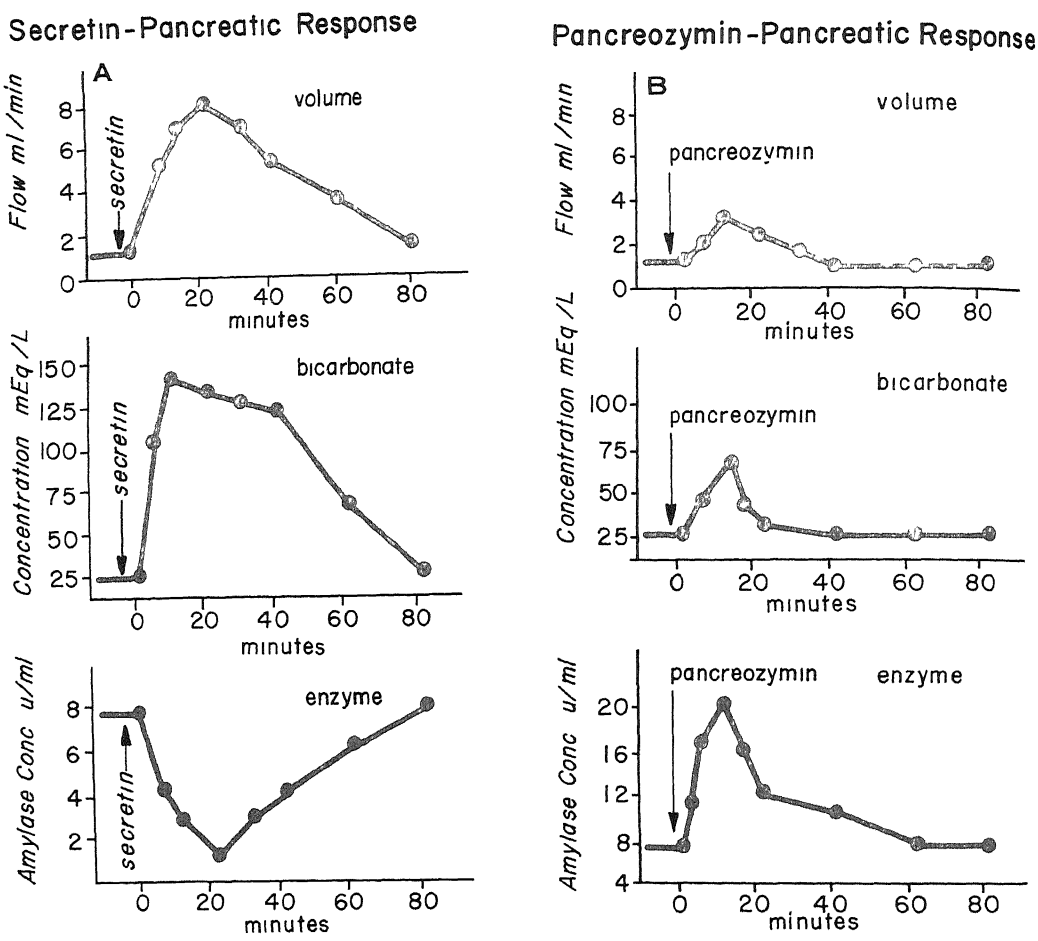


Fig 2 32 Volume bicarbonate and amylase response following the administration of secretin and pancreozymin to a normal subject The hormones were administered intravenously in a dose of 1 unit/kg (Redrawn from Dreiling et al 1964)

usually 30 to 80 min Volume, pH, bicarbonate, and enzyme determinations are made on the samples collected from the duodenum

With normal function, there is a rapid increase in flow rate following secretin injection, with the maximum flow rate usually being reached in 20 min A flow rate of 1 to 2 ml /kg weight/30 min is usually obtained in normal individuals The bicarbonate concentration increases with the flow rate and may go as high as 140 to 150 mEq /liter As the flow rate increases, there is an associated decrease in enzyme concentration In patients with obstruction of the pancreatic ducts due to neoplasm or calculi, there is a reduction in flow rate In diseases of the pancreas associated with destruction of the secreting parenchyma, such as pancreatitis or fibrocystic disease, a reduction in flow rate also is observed In inflammatory diseases of the pancreas, the bicarbonate concentration may be markedly reduced A concentration of 90 mEq /liter is usually taken as the lower limit of normal for bicarbonate

When pancreozymin is used as a stimulus for pancreatic secretion to test pancreatic function, the procedure of the test is the same as that described for the secretin test As a matter of fact, secretin and pancreozymin are frequently used together in testing pancreatic function There is some disagreement as to which hormone should be administered first when they are used together English and European workers have tended to give secretin first, followed by pancreozymin in 30 min (Burton et al , 1960), while workers in the United States seem to prefer the reverse order, giving pan-

creozymin first, followed 10 min. later by secretin (Sun and Shay, 1960). Dreiling and Janowitz (1962) have used both sequences of secretin-pancreozymin and pancreozymin-secretin in testing pancreatic function and have come to the conclusion that the addition of pancreozymin to the standard secretin test technique does not significantly improve the diagnostic accuracy of the procedure.

The separate effects of secretin and pancreozymin on pancreatic secretion of man are illustrated in Figure 2 32.

7 / BILIARY SECRETION

Bile is a mixture of secretory and excretory products consisting of water, bile salts, bile pigments, inorganic salts, and a mixture of lipid material including fats, cholesterol and lecithin. In most animals and man, bile is formed continuously by the liver. In its transport to the intestinal tract, bile first passes along the bile capillaries, then the hepatic and cystic ducts to the gallbladder where it is stored and concentrated. The expulsion of bile from the gallbladder and its passage along the common bile duct into the duodenum is intermittent, related in time to the arrival of food in the intestine, and is quite independent of the formation of bile by the liver.

Hepatic Architecture

Microscopic sections of the normal human liver reveal a rather characteristic arrangement of the hepatic cells. Groups of cells appear to be arranged in a lobular fashion around a *central vein*, with the periphery of the lobule delineated by the presence of *portal triads*. A portal triad consists of a small arteriole of the hepatic artery, a venule derived from the portal vein, and a bile ductule. The central vein of the hepatic lobule drains into the hepatic vein.

It was originally thought that the liver cells were arranged in cords, two cells thick, with a bile capillary embedded in the center of the cord between the cells. It is now realized that liver cells are not arranged in cords as previously supposed, but in the form of *plates* one-cell thick that provide a honeycomb or sponge-like architecture (Elias, 1949, 1953). Throughout the sponge-like structure, the cell plates are tunneled by a communicating system of cavities known as *lacunas*. The lacunas contain the blood sinusoids which have a basement membrane formed of *Kupffer cells*. The very narrow space separating the sinusoidal wall from the liver cell plates is known as *Disse's space*. The sinusoids convey blood from the portal vein and hepatic artery to the hepatic vein.

The *biliary canaliculi* are located within the liver cell plates, and are arranged in a communicating system surrounding the hepatic cells. The liver cell thus forms a narrow wall between the blood stream on the one hand and the bile stream on the other.

The electron microscope has provided new information regarding the relationship of the liver cells to the biliary canaliculi. The bile canaliculus is a space less than 1 μ in diameter between neighboring liver cells. The boundaries of the canaliculus are formed by the plasma membrane of the liver cells, and small finger-like projections or microvilli extend from the liver cells into the lumen of the canaliculus. It will be recalled that this arrangement is characteristic of surfaces across which fluid transfer occurs. In the pericanalicular zone, a membrane system of lamellas and vacuoles is always present. This is the Golgi complex. Free ribosomes, vacuoles, and dark, rounded bodies occur in the pericanalicular zone. It is thought that these structures, as in the pancreatic acinar cell, are related to the secretory functions of the hepatic cells (Schaffner, 1965).

The Composition of Bile

Table 2.9 gives the composition of human liver bile (as parts in 1000). Liver bile has a pH which varies between 8.0 and 8.6. The reaction of human gall bladder bile is neutral or very slightly alkaline, that of the dog (or cat) is definitely acid with a pH between 5.0 and 6.0. The chief biliary components are the *bile salts*, *bile pigments*, *cholesterol*, and *lecithin*. These organic materials make up over 60% of the total biliary solids. As a result of the absorption of water and inorganic salts, gall bladder bile is several times more concentrated in organic solids than liver bile. The biliary constituents may vary independently of one another.

BILE ELECTROLYTES

The electrolyte composition of bile is related to flow rate, in a manner similar to that observed in the pancreas. In turn, flow rate is related to the availability of bile salts. The principal cations in bile are sodium and potassium; the principal anions are bicarbonate and chlo-

ride It has been estimated that 85 to 90% of the bile salts excreted in the bile are reabsorbed and return to the liver for re-excretion

If bile salts are prevented from entering the intestinal tract due to removal through a biliary fistula, biliary flow gradually decreases To overcome this problem, Wheeler and Ramos (1960) infused sodium taurocholate intravenously at a constant rate into dogs which had been prepared with biliary fistulas They were then able to obtain multiple bile samples for measurement of electrolyte concentration without interfering with flow rate They found that the taurocholate concentration of the bile was equal to the difference between the cations (Na^+ and K^+) and the anions (Cl^- and HCO_3^-) This relationship was maintained over a wide range of taurocholate concentrations (40 to 280 mEq /liter)

A characteristic alteration in the bile electrolyte composition was observed with increasing flow rate As the flow rate increased, the taurocholate concentration decreased, indicating that the bile salt was excreted at a constant rate With increase in bile flow, there was a striking increase in bicarbonate concentration and pH, similar to that observed in the pancreas Chloride concentration also increased with increased flow rates The changes observed in chloride and bicarbonate concentration and pH at different flow rates are illustrated in Figure 2 33

The *osmolarity* of the bile samples showed little variation, despite marked differences in electrolyte composition, bile flow, or time of collections In all samples, the osmolarity of the bile was essentially that of plasma The explanation of the apparent discrepancy between varying concentrations of inorganic ions and a relative constant osmolarity is due to the bile salts forming large aggregations or micelles that have low osmotic activity It is the sum of $Na^+ + K^+ + Cl^- + HCO_3^-$ that accounts for the osmolality of both liver and gallbladder bile

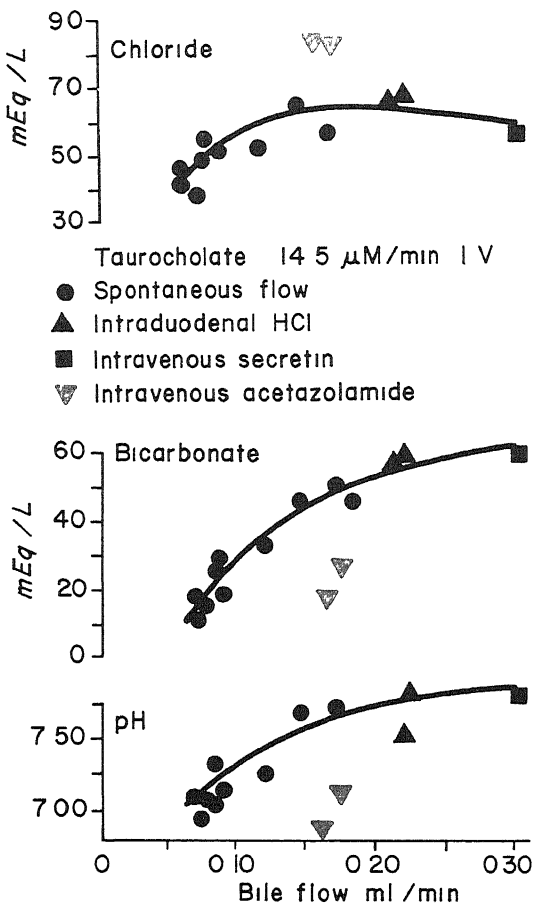


Fig 2 33 Relationship between bile flow and composition Sodium taurocholate was administered by a constant intravenous infusion at a rate of 14 5 micromoles per minute Maximal flow pH and bicarbonate concentration were observed after intravenous secretin or intraduodenal HCl After intravenous acetazolamide the chloride concentration was higher and the bicarbonate concentration lower than at comparable spontaneous lows (Redrawn from Wheeler and Ramos 1960)

Table 2 9
Constituents of human hepatic bile

Substance	Parts/1000
Water	976 22
Solids	23 78
Mucin and pigments	5 00
Bile salts	9 00
Fatty acids from soaps	1 23
Cholesterol	0 63
Lecithin / Fat	0 60
Inorganic salts	7 32

BILE SALTS

The most important secretory products of the liver are the bile acids which appear in the bile as salts of sodium or potassium There are a variety of bile acids to be found in the bile of different species, all of which are formed by conjugation of an amino acid with a bile acid The conjugated bile acid then combines with sodium or potassium to form a bile salt

Bile salts are divided into *primary* and *secondary* types. Primary bile salts are formed by the cells of the liver and secreted into the bile. Secondary bile salts are formed in the intestinal tract as a result of the action of bacteria on the primary bile salts. The secondary bile salts are absorbed by the gut, recirculated to the liver, and then secreted in the bile.

Origin and Formation. In man and other primates, primary bile acids are formed from cholesterol by the action of liver enzymes. In most mammalian species, the main primary bile acids are cholic acid and chenodeoxycholic acid. These primary bile acids are conjugated with the amino acids taurine or glycine.

In the lower part of the small intestine, the primary bile acids are subjected to the action of various intestinal microorganisms which result in the conjugated bile acids being hydrolyzed to free bile acids. Additionally, the 7 α-hydroxyl group is removed from cholic acid and chenodeoxycholic acid yielding deoxycholic acid and lithocholic acid, respectively.

A number of organisms have been identified in the small intestine including *enterococcus*, *lactobacillus*, *bacteroides*, and *clostridium*, which are capable of converting the primary bile acids to secondary bile acids. Figure 2-34 illustrates the major bile salts found in man. Many of the bile acid metabolites resulting from the action of bacteria are not absorbed and some, as lithocholic acid, are absorbed very poorly. Most of the lithocholic acid formed in the intestine becomes tightly bound to intestinal microorganisms and with other bile acid metabolites is excreted in the feces. The main bile acids found in portal blood are cholic acid, chenodeoxycholic acid, deoxycholic acid, and a very small amount of lithocholic acid.

Conjugation of Bile Acids. The conjugation of bile acids with taurine and glycine is catalyzed mainly by the microsomal faction and requires ATP, CoA, and Mg.

The ratio of glycine to taurine conjugates in human bile varies widely due to the fact that the ratio is influenced by diet, hormones, disease, and the types of microorganisms in the small and large bowel.

Control of Bile Acid Formation. The exact mechanisms that control bile acid formation have not been completely defined. Most of our information has been obtained by studying the hepatic metabolism of cholesterol which is the precursor of bile acids. Additional information has been obtained from studies of the rate of bile acid synthesis in animals with biliary fistulas. If bile is continuously drained from a bile fistula, the synthesis of bile acids increases 10 to 15-fold. Infusions of bile acids into the small bowel in such a preparation reduce the rate of bile acid synthesis.

An increase in cholesterol synthesis precedes the increase in bile acid synthesis. Myant and Eder (1961) have found that labeled acetate is

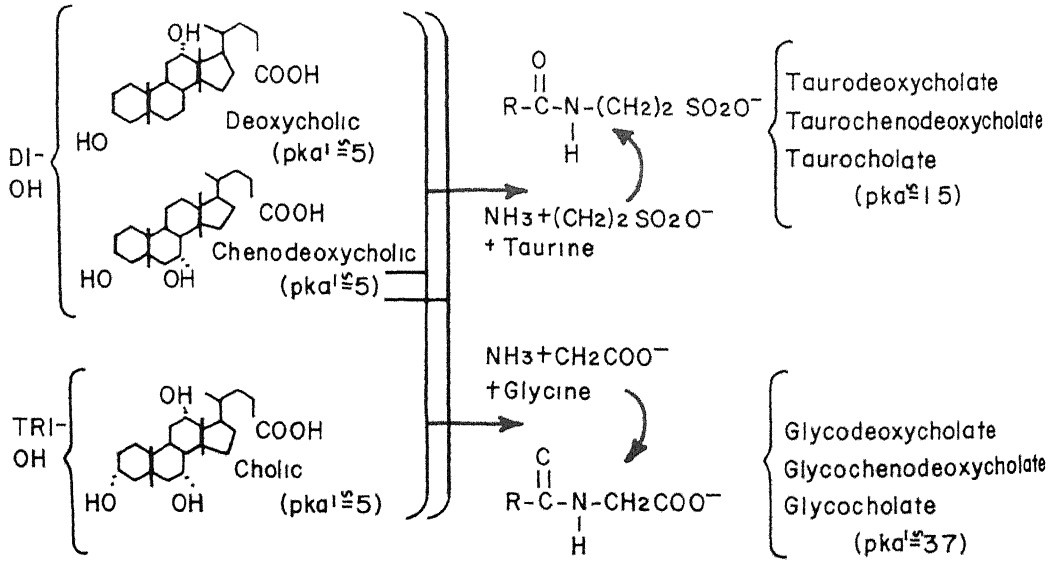


Fig 2-34 Predominant bile salts of human bile. The two primary bile acids—cholic and chenodeoxycholic—and the secondary bile acid—deoxycholic—are conjugated with glycine or taurine in a peptide bond resistant to all digestive enzymes. (From Hofmann, 1968a.)

incorporated into cholesterol at an increased rate in bile fistula rats Beher and associates (1962) have suggested that the synthesis of bile acids is regulated by a "double feedback" mechanism Bile acids inhibit the conversion of cholesterol to bile salts, this leads to an increase in liver cholesterol, and this, in turn, leads to a decreased cholesterol synthesis Bergstrom (1959) has suggested that the site of control of bile acid biosynthesis is the first step in the conversion of cholesterol to bile acids, the 7 α-hydroxylation Recent observations by Danielsson and his associates (1967) support Bergstrom's view for they have found that the rate of 7 α-hydroxylation of cholesterol is several times faster in homogenates from bile fistula rats than in homogenates from control rats, whereas other reactions occur at about the same rate in both control and bile fistula rats

Danielsson (1963) has made an extensive review of the pronounced effect of diet on formation and metabolism of bile acids, including the effects of carbohydrates, steroids, bile acids, vitamins and minerals When the regular diet of man is changed to a synthetic diet, the half-life of bile acids is increased and bile acid production is decreased If indigestible material is added to the diet, bile acid production is increased and the half-life of bile acids is decreased

The state of thyroid activity also influences a rate of synthesis of cholesterol as well as bile acids Hyperthyroid states increase the rate of synthesis of cholesterol as well as that of bile acid production Hypothyroid states are associated with decreased cholesterol synthesis, a prolonged half-life of bile acids and decreased bile acid formation

Bile Salts Micelles The formation of bile salts from cholesterol, which is water insoluble, yields a C-24 carboxylic acid which has a steroid nucleus containing hydroxyl groups The bile acids are polar molecules with high aqueous solubility

Bile salts are similar to ionic detergents and are amphipathic molecules which have both hydrophilic and hydrophobic regions It is these water soluble and fat soluble regions of the molecule which account for the characteristic physical properties of bile salts At very low concentrations, bile salts form molecular solutions When the concentration reaches a critical level, however, the amphipathic molecules form molecular aggregations called *micelles* The concentration at which molecular aggregation occurs is called the *critical micellar concentration*

The micelles formed by bile salts are completely different from those formed by ionic detergents The bile salt molecule has a rigid disc-shaped body that is fixed to a small side chain On one side of the molecule are all of the polar hydroxyl groups, and the charged group is at the end of the short side chain Bile salts are more effective in forming micelles in the presence of polar lipid substances as fatty acids and monoglyceride molecules The present concept of the bile salt micelle is illustrated in Figure 2 35

The Enterohepatic Circulation of Bile Salts In the jejunum, the primary function of the bile salts are to solubilize dietary triglycerides by forming a micellar solution of bile salts, fatty acids, and monoglycerides The fatty acid and monoglyceride are rapidly absorbed from the micelle and the bile salts continue down the small bowel where they are absorbed primarily from the terminal ileum

By use of ^{14}C labeled cholic acid, Lindstedt (1957) studied normal subjects and found that

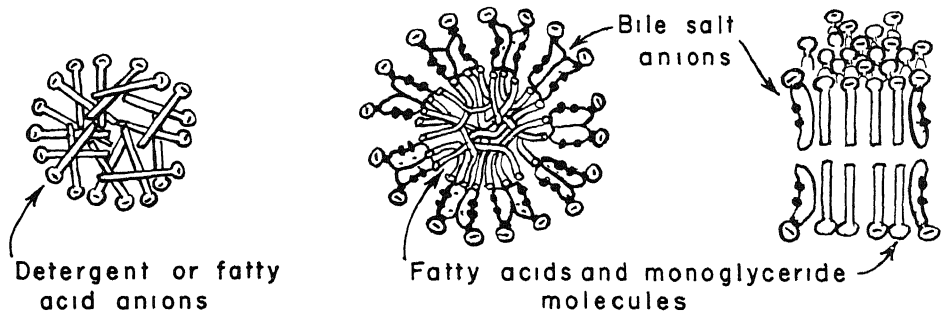
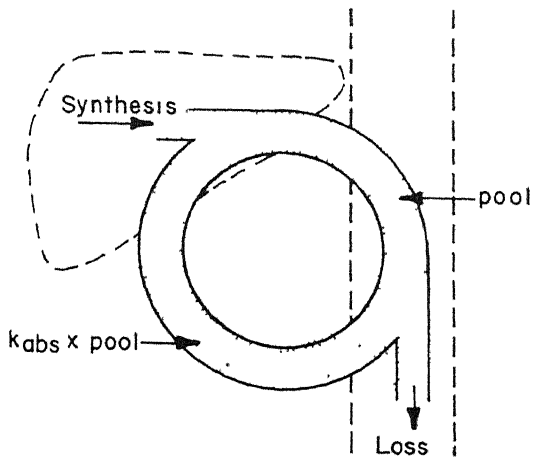


Fig 2 35 The probable molecular arrangement of a typical soap or anionic detergent micelle (*left*) is contrasted to two proposed models for the bile salt-polar lipid micelle (*center* and *right*) A cross section of each micelle is shown the aggregate on the *left* and in the *center* is spherical that on the *right* disc-shaped (From Hofmann 1968a)

the radioactive bile acid was diluted in a large bile acid "pool." The size of this bile acid pool varied from about 3000 to 4000 mg. The half-life of the pool was calculated to be about 3 days. The daily synthesis of bile acids was about 350 mg. Thus the bile acid pool is 6 to 10 times the amount synthesized daily.

In the lower part of the ileum, the primary bile salts, as well as secondary bile salts resulting from microbial action, are absorbed and returned to the liver via the portal blood. They are then resecreted into the bile and delivered back to the small bowel. This secretion, absorption, and resecretion is known as the *enterohepatic circulation* of bile salts. It is estimated that each bile salt molecule participates in 15 to 20 enterohepatic cycles before excretion. Figure 2-36 illustrates the enterohepatic circulation of bile salts.

The bile salt pool remains relatively constant as synthesis is balanced against excretion. Synthesis is under the influence of the amount of bile salts absorbed in the terminal ileum, approximately 80 to 90% of all bile salts secreted in the bile are reabsorbed by the small bowel. In man 20 to 30 g of bile acids are secreted by the liver daily in the bile; however, only 0.5 to 0.7 g of bile acids appear in the feces each day.



$$k_{abs} = \frac{\text{bile acid absorbed/cycle}}{\text{bile acid pool}}$$

Fig. 2-36 Schematic representation of the enterohepatic circulation of bile salts. It is customary, in the manner of Lindstedt (1957), to estimate the size of the bile salt pool by extrapolating the specific activity decay curve for biliary bile acids to zero time. The bile salt pool thus represents the sum of the fraction of the bile salt pool absorbed during the previous cycle (*K* pool) plus that bile acid synthesized by the liver during that particular enterohepatic cycle. If the pool remains constant in size, daily loss is equal to $(1 - K)\,PC$, where C is the number of enterohepatic cycles per day (Hofmann, 1966). Most of the bile salt reabsorption is considered to occur in the ileum where there is an active transport mechanism for bile salts. (From Hofmann, 1968a.)

Excretion of Bile Salts. The main pathway for excretion of bile salts is with the feces. Less than 5% of bile acids excreted appear in the urine. Under normal conditions, bile acids found in feces are a complex mixture of compounds most of which have been formed by microbial action on the primary bile salts rendering them essentially unabsorbable. The complexity of the mixture of fecal bile salts makes their quantitative determination difficult and of little clinical importance.

THE BILE PIGMENTS

Bilirubin ($C_{33}H_{36}N_4O_6$) is the chief pigment in human bile and in the bile of the carnivora. Biliverdin ($C_{33}H_{36}N_4O_8$) is an oxidative derivative of bilirubin, and although it is present in only small amounts of human bile, it is the chief pigment of birds' bile. Besides biliverdin, other derivatives of bilirubin are found in the body. *Urobilinogen* is a reduction product of bilirubin, upon oxidation it yields *urobilin*. *Bilifuscin* and *bilifuscin* are formed by the oxidation of biliverdin. The latter two pigments are not found in bile but may be present in gall stones. The pigments constitute from 1.5 to 2.0% of the total solids in liver bile.

Origin of the Bile Pigments. Bile pigments are derived from the free hemoglobin resulting from the destruction of red cells in the body, and in smaller amounts, from other chromoproteins. Approximately 85% of bilirubin is derived from senescent red cells. The red cells have a life span of about 100 to 120 days. Those that perish are disposed of by the reticuloendothelial cells distributed throughout the body. From these cells an average of about 8 g of hemoglobin is released per day, this yields about 300 mg of bile pigments. In order to produce bile pigment from hemoglobin, the hemin portion must be separated from the protein (globin) portion of the hemoglobin molecule. Oxidation scission of the iron-porphyrin follows. This involves opening the tetrapyrrole ring at the alpha position and the "shelling out" of iron. The resulting iron-free pigment is probably biliverdin, which is later reduced to bilirubin. All these reactions are presumed to take place in the reticuloendothelial cells, but

the order in which they occur is not known In particular, it is not known if the splitting of the hemin moiety on the protein portion of the molecule must take place before the porphyrin ring is open

Bilirubin derived from hemoglobin is unconjugated and known as "free bilirubin" It is highly lipid soluble and has limited aqueous solubility Free bilirubin is bound by albumin which greatly increases its solubility in plasma Each mole of human albumin is capable of binding at least 2 moles of bilirubin Other plasma proteins, such as α1- and α2-, and β-globulins, bind bilirubin to a limited extent when plasma bilirubin concentrations exceed 20 to 30 mg /100 ml

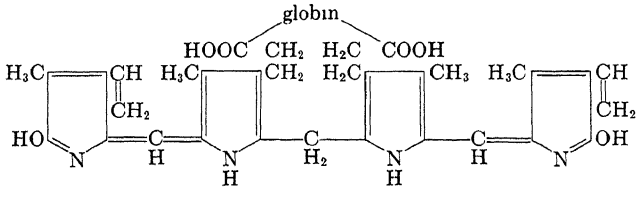
The binding of free bilirubin by albumin is of considerable clinical importance Bilirubin encephalopathy or kernicterus which occurs in infants shortly after birth is due to high concentrations of free bilirubin in the plasma The highly lipid soluble unconjugated bilirubin enters the neurons of the basal ganglions, hippocampus, cerebellum and medulla causing necrosis of these nerve cells Adequate available binding sites on albumin prevent this from occurring Anionic drugs as sulfonamides are also bound by albumin and compete with bilirubin for the available binding sites Thus, the administration of sulfonamides to an infant with unconjugated hyperbilirubinemia displaces bilirubin from albumin increasing the risk of kernicterus Free fatty acids produce a similar effect and asphyxia, hypoxia, and other conditions leading to acidosis also increase the risk of kernicterus

The chemical relationship between bilirubinglobin and the various bile pigment derivatives is shown schematically in Diagram 1

Bile pigments may be formed, not only in the liver, but also in the spleen and bone marrow and at various other places in the body This was not generally recognized until the experiments of Whipple and Hooper (1913) and the more conclusive experiments of Mann and his associates (1924) The former attempted to isolate the liver from the general circulation, whereas the latter removed the liver completely In either case, bile pigments continued to be formed, and, since they could not be excreted through the liver, they accumulated in the blood and produced a condition analogous to jaundice Mann and his associates also showed that the blood of the splenic vein of a normal animal has a higher bilirubin concentration than does the blood in the corresponding artery

By use of ^{15}N-labeled glycine, in vivo, the non-hemoglobin origin of bilirubin has been better defined When ^{15}N-labeled glycine is administered, the label appears incorporated in stercobilin within a few days This rapid appearance of the isotope in the intestinal oxidation product of bilirubin has been termed "early labeled bilirubin" When isotopically labeled δ-aminolevulenic acid is administered, the radioactive label is more readily incorporated into early labeled bile pigment than into hemoglobin heme. The principle sites of formation of "early labeled bilirubin" are the liver and bone marrow Most nonhemoglobin heme proteins, such as myoglobin, catalase and cytochromes, are degraded in the liver and these proteins contribute to the early labeled bile pigment

Possible sources of bone marrow origin include the intracorpuscular degradation of hemoglobin during erythrocyte maturation and destruction of newly formed erythrocytes in the bone marrow before they reach the circulation



Bilirubinglobin

Bilirubinglobin

hydrolysis —globin

biliverdin $\xrightarrow{\text{reduction}}$ bilirubin $\xrightarrow{\text{reduction}}$ urobilinogen $\xrightarrow{\text{oxidation}}$ urobilin

Diagram 1

In patients with pernicious anemia sickle cell anemia and congenital porphyria the early labeled pigment is responsible for a major portion of the bile pigment excreted Early labeled pigment is also found in patients with disorders of erythropoieses as thalassemia and after hemorrhage

Hepatic Uptake of Bilirubin The unconjugated bilirubin formed in the reticuloendothelial cells is transported to the liver tightly bound to albumin The exact mechanism of uptake of the unconjugated bilirubin by the hepatic cell is unknown However it is known that unconjugated bilirubin in the plasma enters the liver preferentially with respects to other tissues

It has been suggested by Gartner and Arias (1969) that two possibilities exist for the rapid uptake of unconjugated bilirubin One is that the relative specificity of the hepatic cell to remove unconjugated bilirubin from the plasma could reside in the plasma membrane facing the sinusoid Lateral extensions of the membrane in this area may have specific molecular configurations that permit the entrance of the unconjugated bilirubin into the cell This suggestion is based upon the heterogeneity of enzymatic activity in different portions of the hepatic cell plasma membrane and the binding of sulfobromophthalein (BSP) and other organic anions by liver plasma membranes Equal molar amounts of organic anions as taurocholate, unconjugated bilirubin, and fluorescein, which do not compete with BSP uptake when simultaneously injected in vivo, show little or no interference with BSP binding by plasma membranes in vitro Other organic anions, however, as indocyanine green, which compete with BSP for hepatic uptake in vivo, considerably reduces BSP binding by isolated plasma membranes in vitro These findings suggested to Gartner and Arias that the plasma membrane of the liver cell is not molecularly homogenous and that different specific receptor sites for various organic anions may be located in the lateral extensions of the plasma membranes facing the sinusoid

Another possibility is that the unconjugated bilirubin may be dissociated from albumin at the plasma membrane and be transferred to an intracellular protein which acts as an acceptor and facilitates the transfer of bilirubin into the liver cell Recently, two basic albumin proteins designated as Y and Z have been isolated from the cytoplasm of hepatocytes and they account for most of the intracellular binding of bilirubin and BSP Apparently, Y has a greater affinity for bilirubin and BSP than Z Binding on Z does not occur until concentration of bilirubin and BSP on Y reaches a critical level

Conjugation of Bilirubin Within the liver cell bilirubin is conjugated with glycuronic acid forming bilirubin glycuronide (Schmid, 1956, Billing et al 1957) The conjugation of bilirubin is accomplished by a microsomal enzyme glucuronyl transferase which utilizes uridine diphosphoglucuronic acid as a donor The lipid soluble unconjugated bilirubin is thus converted to a conjugated water soluble form Since bilirubin is capable of combining with two molecules of glucuronic acid, either the mono- or diglucuronide may be formed Glucuronyl transferase is also found in skin, ovary, adrenal glands, kidneys, testes, and synovial membrane Conjugation may take place to some extent in these areas, but the greatest activity of this enzyme occurs in the liver where it is associated with the endoplasmic reticulum membrane portion of the parenchymal liver cell It has been suggested that the extrahepatic glucuronyl transferase results in "pigment I" or bilirubin monoglucuronide and then converted to the diglucuronide, "pigment II," in the liver Bilirubin excreted in the bile is essentially bilirubin diglucuronide The existence of the monoglucuronide in bile is doubtful

A number of clinical conditions are now recognized as being due to abnormalities of glucuronyl transferase activity. Several drugs, dyes and steroids may inhibit glucuronyl transferase activity A small percentage of nursing mothers excrete in their breast milk an unusual isomer of pregnanediol, which markedly inhibits the glucuronyl transferase activity of their infant This leads to prolonged benign unconjugated hyperbilirubinemia until about the 12th week of life unless breast feeding is terminated These mothers apparently have an inheritable defect in steroid metabolism that affects the course of hyperbilirubinemia in their neonates

Chronic, nonhemolytic, unconjugated hyperbilirubinemia in man and animal has long been recognized and it is now known that glucuronyl transferase deficiency or absence can account for these conditions The Gunn rat, a mutant Wistar rat, has a congenital absence of glucuronyl transferase These animals have a lifelong, chronic, nonhemolytic, unconjugated

hyperbilirubinemia They are capable, however, of excreting administered conjugated bilirubin Kernicterus frequently develops in these animals

Crigler and Najjar (1952) first described a similar type of nonhemolytic acholuric jaundice developing in infants shortly after birth These infants' jaundice was found to be due to high levels of unconjugated bilirubin in the serum An absence or marked deficiency of glucuronyl transferase in the liver tissue of these infants has been demonstrated Although rare, a congenital absence of the conjugating enzyme, glucuronyl transferase, may be a cause of jaundice in man, as previously shown to be the case in the Gunn rat

In the so-called *physiological jaundice* that develops in most full-term infants between the 2nd and 7th days of life, there is no evidence of hemolysis and the jaundice is of the acholuric type In newborn humans, a deficiency of glucuronyl transferase activity has been demonstrated The activity of the adult liver is not reached until about the 10th day of life These observations suggest that physiological jaundice of the newborn is related to a delay in the development of the glucuronide-conjugating system In premature infants, the deficiency in glucuronyl transferase activity may lead to severe jaundice and kernicterus

Phenobarbital is one of a number of drugs that can increase glucuronyl transferase activity This discovery has led to the treatment with phenobarbital of some patients with the Crigler-Najjar syndrome who have mild unconjugated hyperbilirubinemia In these patients, who have only a deficiency of glucuronyl transferase, there is a dramatic amelioration of jaundice on administration of phenobarbital Phenobarbital has also been administered to women before delivery and it has been shown that this will reduce serum bilirubin concentrations in their infants by approximately 50%

Hepatic Excretion of Bilirubin. Bilirubin must be conjugated with glucuronic acid prior to excretion The mammalian liver has a large reserve in regard to excretory capacity for handling bilirubin However, the process is rate limiting In adult rats, it has been determined that the maximum excretory rate is approximately 60 μg of bilirubin excreted per 100 g /min The high ratio of bile bilirubin to plasma suggests that the process is energy dependent and the excretion of bilirubin is an example of active transport The exact mechanism, however, is not known The microvilli, extending into the biliary canaliculi, and the presence of several phosphatases associated with the plasma membrane, also suggest that the process is an active one

When severe hemolysis occurs in patients with normal liver function, the degree of unconjugated hyperbilirubinemia is slight In the presence of liver disease, however, jaundice may be marked Conjugated hyperbilirubinemia occurs only in the presence of liver disease The capacity of the human liver to handle conjugated bilirubin is not known In the rat, administered conjugated bilirubin is cleared about as rapidly as administered unconjugated bilirubin

The Dubin-Johnson syndrome (1954), which occurs in man, is an example of a defect in the excretion of conjugated bilirubin A similar condition has been found in sheep The liver has a characteristic black color which is due to retention of a pigment, originally thought to be lipofuscin, in lysosomes of the hepatic cells Recent studies suggest that the pigment is a melanin and possibly results from impaired biliary excretion of epinephrine metabolites

In the mutant sheep, with an excretory defect comparable to the Dubin-Johnson syndrome in man, it has been found that the excretion of administered taurocholate is apparently normal This suggests that bile acids and bile pigments are excreted by the liver cell via different pathways

In 1948, Rotor and his colleagues described cases in which there was chronic conjugated hyperbilirubinemia but no abnormal pigment was present in the livers of these patients This condition, thought to be transmitted by an autosomal dominant gene, is also considered as a clinical example of an excretory defect in the excretion of bilirubin, BSP, and other organic anions In both the Dubin-Johnson syndrome and the Rotor syndrome, the excretion of BSP is essentially zero However, hepatic storage of BSP is apparently normal Conjugation of bilirubin with glucuronide apparently proceeds in a normal fashion, but excretion does not occur for some unknown reason

Within the past few years, it has been discovered that various drugs, such as anabolic steroids, estrogens, and oral contraceptive agents, reduce the capacity of the liver to excrete organic anions such as BSP During pregnancy,

some women manifest chemical and morphological signs of cholestasis during the last trimester Additionally, when they take oral contraceptive agents, they may also have mild jaundice It is thought that these patients must have some defect in the excretory mechanism for handling bilirubin and BSP With the further depression of this mechanism by the estrogens of pregnancy or oral contraceptives, they develop cholestasis

Enterohepatic Circulation of Bile Pigments The bilirubin that enters the intestine undergoes reduction by bacteria to form *urobilinogen* (also called *stercobilinogen*) A part of this is excreted in the stools and by exposure to air, is oxidized to *urobilin* Some of the urobilinogen is reabsorbed into the portal circulation and passes to the liver where it is reconverted in part to bilirubin in the bile (Figure 2 37) It is then almost entirely re-excreted both as bilirubin and as urobilinogen Any urobilinogen which escapes into the general circulation and is excreted by the kidney becomes oxidized to *urobilin* after the urine has been voided Normally, however, no urobilin, or mere traces (0 5 to 2 0 mg), appears in the urine Only traces are present in normal blood Bilirubin itself does not normally appear in the urine The urobilinogen normally present in bile merely represents re-excretion of this pigment after its absorption from the intestine When the entire output of bile is collected through a biliary fistula, there is complete disappearance of urobilinogen from the bile after the pigment already present in the intestine has been carried away in the feces When a part of the bile is allowed to enter the intestine, or bile is fed by mouth the derived pigment invariably appears in the fistula bile

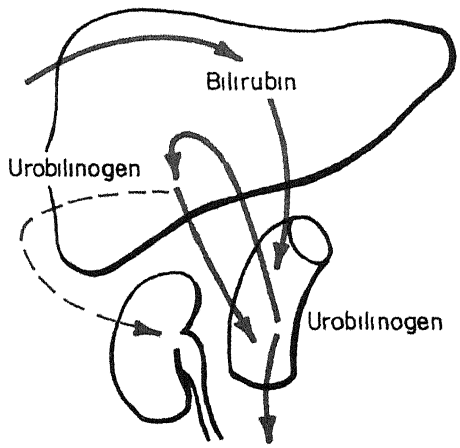


Fig 2 37 Diagram illustrating the formation of urobilinogen from bilirubin in the intestine, its excretion in part in the feces and its absorption in part into the portal blood. Normally, the absorbed pigment except in negligible amounts is practically all re-excreted in the bile The dotted lines indicate its passage into the blood, and its excretion by the kidney in cases of liver damage, excessive blood destruction, or in the early stages of obstructive jaundice

Complete experimental obstruction of the common bile duct, since it prevented the re excretion of the urobilinogen absorbed from the intestine, also resulted in its accumulation in the blood and its excretion in the urine These effects, however, can only occur for a short time after the duct has been obstructed and are due to the absorption of pigment remaining in the intestine from the period prior to obstruction After the pigment has been cleared from the intestine, the urine, although it may contain large amounts of bilirubin, is quite free from the derived pigment Depression of the excretory function of the liver by hepatic poisons or disease causes urobilinogen to appear in the urine When bilirubin formation is increased by hemolytic agents, urobilinogen production is also increased, the liver, even though its function is normal, is unable to re-excrete the excess pigment absorbed from the intestine, and urobilin appears in the urine

The availability of unconjugated and conjugated bilirubin ^{14}C has made it possible to study the enterohepatic circulation of bile pigments in a more precise manner Lester and Schmid (1963a) administered both conjugated and unconjugated bilirubin ^{14}C to rats which had been prepared with an external biliary fistula These investigators found that both of the labeled pigments were absorbed and appeared promptly in the bile The urinary excretion of the isotope was less than 5% of the dose Because essentially all of the labeled pigment re-excreted in the bile was conjugated, it could not be decided whether the unconjugated and conjugated bilirubin had been absorbed intact or had undergone chemical modification during the absorptive process The experiments were repeated in a strain of Wistar-Gunn rats that have a congenital deficiency of hepatic glucuronyl transferase Although they are unable to conjugate bilirubin, they are able to excrete conjugated bilirubin Thus, if conjugated bilirubin was unaltered in the absorptive process, it should be excreted in the bile of the Gunn

rats It was found, after intraduodenal administration of either unconjugated or conjugated bilirubin ^{14}C, that no radioactive pigment appeared in the bile, indicating that conjugated bilirubin had been altered in the absorptive process

The absorption of urobilinogen ^{14}C has also been studied (Lester and Schmid, 1964) In rats, it has been found that significant quantities of intact urobilinogen is absorbed from the intestine and re-excreted predominantly in the bile

In man, Lester and Schmid (1963b) have demonstrated that unconjugated bilirubin is readily absorbed from the intestine but little conjugated bilirubin is absorbed (Figure 2 38) Evidence indicates that conjugated bilirubin remains intact during its transit through the biliary tree and small bowel Conjugation prevents reabsorption of the pigment, thus the enterohepatic circulation of bilirubin is minimal

When bacteria in the terminal ileum and colon reduce bilirubin to urobilinogen, it is not known if the glucuronide ester remains intact during these reductive steps Conjugated forms of urobilinogen have not been isolated, but if they do occur, Lester (1964) indicates that minimal absorption of intact urobilinogen glucuronide would be anticipated

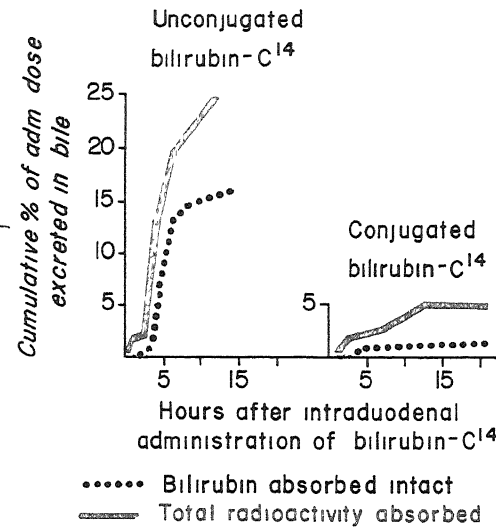


Fig 2 38 Pigment absorption after administration of unconjugated ^{14}C labeled bilirubin and conjugated ^{14}C labeled bilirubin to two human subjects (From Lester and Schmid (1963b))

OTHER CONSTITUENTS OF BILE

Lecithin and Cholesterol. *Lecithin* is present in human liver bile to the extent of from 0 02 to 0 05% The *cholesterol* content is normally from 0 04 to 0 16% The cholesterol is present in the free state, i e , not in the form of esters The percentages of these materials in the bile of the gallbladder are, as a result of the absorption of water and salts through the gallbladder mucosa, much higher than the percentages in liver bile The ratio of the concentration of cholesterol in bile to that of the bile salts is from 1 20 to 1 30 The ability of the bile to hold cholesterol in solution is, in large degree, dependent upon the bile salts If the ratio falls to 1 13, precipitation of cholesterol occurs If bile is dialyzed against water, bile salts are removed and the bile becomes turbid, as a result of cholesterol precipitation

Lecithin and cholesterol are kept in solution by the bile salt micelles In the intestinal tract, the bile cholesterol is mixed with dietary cholesterol and is absorbed principally in the upper small bowel Of the total amount of cholesterol delivered to the small bowel from the bile and the diet, only about 50% is absorbed Most of the unabsorbed cholesterol is converted to coprostanol, a neutral sterol, by microorganisms in the large intestine In germ-free animals, or animals treated with antibiotics, dietary and biliary cholesterol appears in feces in its free form

Bile Proteins. Immunologic techniques have demonstrated the presence of normal plasma proteins in human bile Rawson (1962) has identified albumin and γ-globulin in human bile obtained at autopsy, and from the fistula drainage of patients who had undergone biliary surgery. Russell and Burnett (1963) have used the Ouchterlony gel diffusion technique, immunoelectrophoresis, and gel electrophoresis for the identification of plasma proteins in human bile Samples of bile were obtained from both the gallbladder and common hepatic duct in patients undergoing laparotomy for various reasons Bile samples were obtained from patients with gallstones as well as from patients who had a normal gallbladder and biliary system No differences between normal bile and that obtained from patients with gallstones were found

The presence of plasma proteins in the bile of patients with inflammatory disease of the gallbladder or biliary system on postmortem would

not be too surprising, as one would expect plasma proteins to escape through the mucuous membrane lining the biliary system and gallbladder The presence of plasma proteins in normal bile is not as easily explained However, the liver synthesizes large quantities of albumin and the presence of albumin in normal bile during life might be anticipated. On the other hand, γ-globulin is not formed in the liver Russell and Burnett point out that the hepatic pericellular space (the space of Disse) communicates freely with the liver sinusoids as well as the biliary canaliculi but that molecules with a molecular weight greater than 15,000 are unable to pass from blood to bile Thus, further investigation is needed to determine the pathway by which plasma proteins gain entry into the bile in the absence of disease

Control of Bile Secretion

Bile is secreted continuously and is an excretion as well as a secretion On a normal feeding schedule, more bile is secreted during the daylight hours than during the night More bile is secreted after meals than in the fasting state, and meals are normally eaten in the daytime. If an animal is fed at regular intervals throughout the 24 hr , as much bile is secreted at night as in the daytime The stimulating effect of food on the volume output of bile is due mainly to its protein content Kocour and Ivy (1937) studied a variety of foods for their effect on the output of bile and found that only meat and liver were consistently effective

Before discussing various agents influencing the flow of bile, it is well to define some of the terms that will be used A *cholagogue* is any agent which increases the flow of bile into the intestine The bile may come from the gallbladder, so the word does not necessarily signify an increase in the secretion of bile by the liver A *choleretic* is an agent which increases the output of bile from the liver without necessarily changing its concentration, thus a cholerectic will usually increase the output of bile solids, as well as of the liquid portion of the bile A *hydrocholerectic* causes an increase in the volume of bile without a corresponding increase in the output of bile solids.

The natural bile salts have a powerful cholerectic action on the liver No doubt a great part of the increase in the flow of bile after a meal is the result of the discharge of bile from the gallbladder, which increases the amount of bile salts in the intestine and consequently in the portal blood. Certain oxidized bile acids, notably dehydrocholic acid, act as hydrocholerectics and are sometimes used in therapy to produce an abundant flow of dilute bile

Secretin increases the secretion of bile by the liver although it does not contract the gallbladder Studies by Grossman and his coworkers (1949) indicate that secretin acts as a hydrocholerectic in man The site of action of secretin is thought to be the small bile ducts or ductules Its effect results in increasing the volume of bile by addition of water and markedly increasing the bicarbonate concentration The onset of effect is similar to that of the pancreas Some recent evidence suggests that an electrolyte fraction may be secreted which is dependent on the active transport of Na and is inhibited by inhibition of ATPase

Secretin causes only a moderate increase in bile flow, and none at all if the liver is already secreting vigorously The maximal rate of flow that can be induced by secretin is only about one-tenth the rate that can be attained by the administration of bile acids and this ratio cannot be increased by increasing the dose of secretin These facts suggest that secretin is not a major factor in the regulation of bile secretion.

In addition to humoral and hormonal factors affecting the secretion of bile, the autonomic nervous system may take part in the regulation of bile flow The precise roles the autonomic nervous system play in bile secretion have not been clearly defined

In 1938, Tanturi and Ivy demonstrated that vagus nerve stimulation produced a marked increase in bile flow in dogs More recently, work from Brooks' laboratory (Fritz and Brooks, 1963, Jones *et al* , 1965) has shown that both feeding and insulin hypoglycemia causes a choleresis in dogs However, this can be prevented by vagotomy or anticholinergic drug administration They also demonstrated that the gastric antrum was necessary for the vagal effect on bile flow This suggested that the true effect was probably caused by gastrin Additional evidence by Zaterka and Grossman (1966) also indicate that gastrin has a cholerectic effect Thus, a direct vagal effect on bile flow has not yet been demonstrated

The diagram shown in Figure 2 39, taken from Wheeler, (1968) illustrates some of the factors known to affect bile flow.

Denervating the liver, insofar as this is possible, causes an increase in bile flow, distention of the colon or stimulation of any one of a number of visceral afferent nerves causes inhibition of the flow of bile from the liver It is not known whether the inhibitor effect is due to direct inhibition of the liver cells or to alterations in hepatic blood flow

Mechanism of Bile Secretion

In considering the mechanism of bile secretion, it is convenient to divide the bile constituents into two classes (Brauer, 1958) Class A, which includes water, the inorganic salts, and other highly diffusable constituents, and Class B, which includes the bile pigments, bile acids, and certain foreign substances such as Bromsulphalein, which may be used in liver function tests and appear in the bile

Pure mechanical filtration is ruled out, for bile may be secreted under a pressure which is higher than the pressure of the blood in contact with the liver cells As pointed out by Brauer, in order for a filtration process such as is utilized in the kidney to be effective, it would be necessary for the capillary pressure to exceed the maximum pressure of bile in the bile ducts by an amount at least equal to the colloid osmotic pressure of the blood, this condition is not attained in the liver The flow of bile has been shown to be independent of blood flow or of blood pressure, so long as the oxygen supply is adequate. Another bit of evidence is that cooling the liver greatly reduces the rate of flow of bile For example, in the rat liver, a 10° C reduction in the perfusion temperature between 18° and 38° C reduces bile flow by a factor of 4 Such a change in temperature should have little or no effect on a pure mechanical process. Ether anesthesia also reduces the flow of bile The situation is not so clear with respect to the inorganic salts, and it may be supposed that they are transformed along with the water The

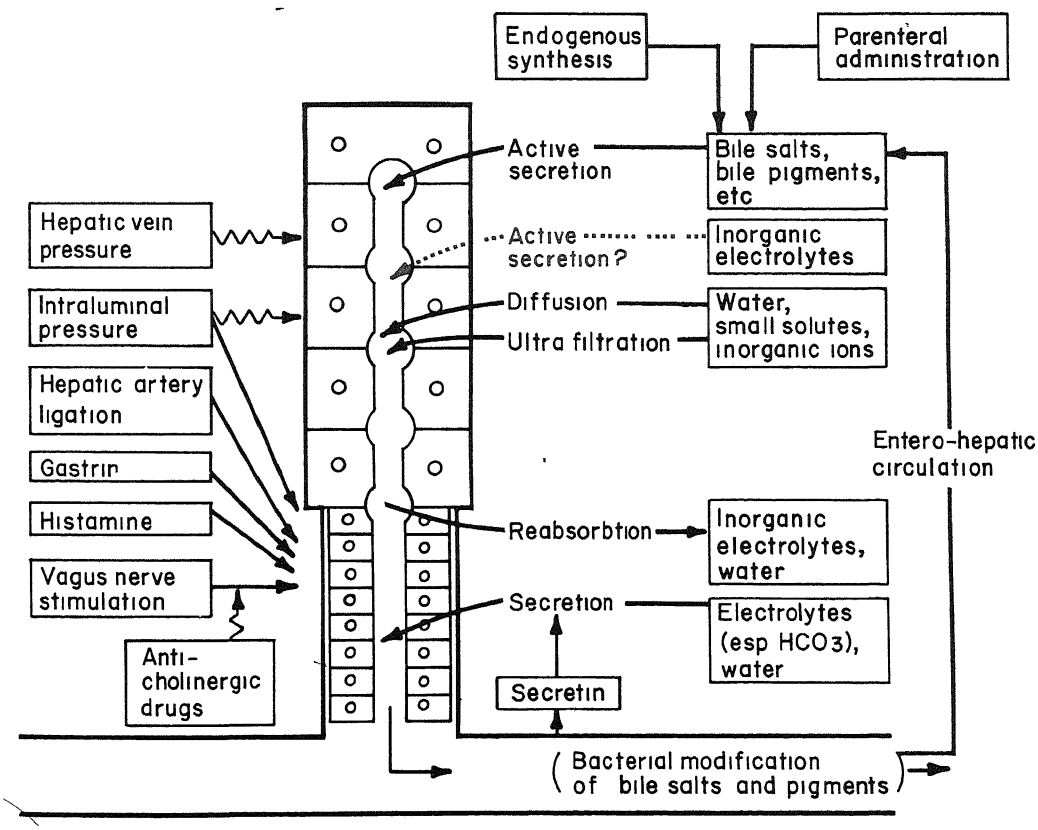


Fig 2 39 Diagrammatic summary of some of the factors known to affect bile flow and electrolyte composition Mechanisms shown on the right are believed to act at canalicular or duct level as shown Locus and mode of action of factors shown on the left is speculative (From Wheeler 1968)

evidence suggests that there is a free interchange of inorganic salts between the blood and the bile as the bile passes through the bile ducts, thus osmotic equilibrium is established between the blood and the bile.

It is interesting to note that, if the concentration of salts in blood plasma is increased so as to make the blood hypertonic, vacuoles appear in the polygonal cells of the liver. This is taken as evidence that these cells are involved in secretion of water into the bile. These hepatic cells secrete the water, as is further evidenced by the fact that in certain forms of toxic hepatitis in which only the polygonal cells are damaged, a sharp reduction in the volume of bile is one of the earliest signs (Brauer, 1958). Bile may be reabsorbed into the blood if the pressure of bile in the bile ducts is elevated above a certain critical level, but at normal pressures reabsorption of bile is negligible.

Organic constituents of the bile are actively secreted. They appear in such concentrations in the bile relative to their concentration in the blood that an active process is inevitable. The liver cells exhibit a certain maximum capacity for secreting various of the bile constituents, comparable to the tubular maximum of the kidney tubules for certain of the constituents of the urine. For example, there is a maximum rate of secretion of Bromsulphalein into the bile beyond which an increase in the dose, or in the rate of infusion in the blood, no longer alters the secretion rate. Similar phenomena have been observed for sodium dehydrocholate, bilirubin, and certain dyes. The same cells are involved in the secretion of different organic compounds, as is indicated by the fact that the secretion of some of these by the liver cells limits the rate of secretion of certain others. For example, if the liver is actively secreting Bromsulphalein, its capacity to secrete bilirubin is limited, and if it is actively secreting bilirubin at a rate near its maximum capacity, it is unable to secrete bile salts at their normal maximal rate. This indicates that there is competition within the individual liver cells among the various substances that are secreted into the bile.

It has been possible to observe, in suitable preparations, the path taken by certain of the substances that are secreted into the bile. For example, fluorescein can be seen to appear first in the parenchymal cells of the liver, then secreted into the bile canaliculi, and then transported down the bile ducts. Similar observations are possible by different methods using radioactive isotopes. Thus it appears that all the constituents of the bile, both organic and inorganic, are normally secreted by the liver cells and by a process which involves the expenditure of energy.

The Gallbladder and Bile Ducts

ANATOMY

The human gallbladder has a capacity of about 50 ml. Its wall is composed of a thin layer of muscle fibers and fibroelastic tissue, with a lining of mucous membrane. The muscle fibers are sparse, and loosely interlaced with one another and with the strands of fibroelastic tissue. The mucosa is surmounted by a layer of columnar epithelium.

The *cystic duct*, through which the bile enters and later leaves the viscus, is tortuous, or S-shaped, and shows spiral folds of mucosa—the *valves of Heister* within its lumen. These so-called valves do not have a valve-like action, for they offer little resistance to the passage of bile in either direction, nor do they seem to prevent a too rapid flow of bile as some have supposed. They develop late in phylogenetic history, apparently being associated with the erect posture and they are found only in primates. The folds are formed in the embryo by the twisting or winding of the duct during development. Their function, as suggested by Keith (1903) and supported by Lichtenstein and Ivy (1937), is to stiffen the wall of the duct and so prevent its kinking.

The *common bile duct*, formed by the union of the hepatic and cystic ducts, passes very obliquely through the muscular wall of the duodenum and joins with the pancreatic duct to form the *ampulla of Vater*.

Though it possesses important functions, the gallbladder is not indispensable, since it can be removed with impunity. After such an operation, however, the larger bile ducts undergo some dilation, which may, in part, compensate for the removal of the viscus. The gallbladder is absent in some animals whose habits and digestive processes are not apparently essentially different from those of animals that possess one. It is absent in the horse, deer, and rat but present in cattle, sheep, dogs, cats, and mice. It is present in fish, amphibia, reptiles, and birds, but is absent from orders lower than these.

Fine Structure of the Gallbladder Mucosa. The mucous membrane of the gallbladder of various animals and man has been studied in detail by light microscopy However, only recently has the fine structure of the gallbladder mucosa of man been studied by electron microscopy The mucous membrane of the gallbladder of man is composed of three distinct types of cells that have been designated "ordinary epithelial cell," "pencil cells," and "basal cells " The most common type of cell is the ordinary epithelial cell, which is arranged as a single layer of tall columnar epithelium The so-called "pencil cell" is rarely seen but appears to be a compressed ordinary epithelial cell Like the ordinary epithelial cell, it extends from the basement membrane to the lumen of the gallbladder The third type of cell, the so-called "basal cell," is scattered randomly throughout the mucous membrane, occupying a position adjacent to the basement membrane The basal cells do not extend to the luminal surface of the mucous membrane

Electron microscopy reveals that the columnar epithelium of the gallbladder has many characteristics of an active absorbing membrane (Evett et al , 1964) The luminal surface of the epithelial cells is composed of a microvillus border The plasma membrane of the microvilli is continuous with that of the intervillus surface and the lateral cell walls of the epithelial cell At the luminal junction of adjacent epithelial cells, the plasma membrane forms "terminal bars" which are thought to prevent intraluminal material from gaining entrance into the intercellular spaces The lateral walls of adjacent cells are interlocked into each other by folds in plasma membranes The basal wall of the ordinary epithelial cell rests on the basement membrane and is straight

The cytoplasm and organelles of the epithelial cells are stratified in a manner similar to that observed in other cells which participate actively in secretory or absorptive functions Just below the microvillus border, there is a small cuticular zone that is clear, due to the absence of mitochondria or granules Immediately beneath this is the subcuticular or dark zone which extends over a third of the cell's height It is in this zone that many mitochondria, vesicles, granules, and a dense endoplasmic reticulum are observed Beneath the subcuticular or dark zone is observed the Golgi apparatus, which appears to be quite similar to that observed in other secreting or absorbing cells The nucleus is large and rounded and occupies the lower third of the cell It possesses no features to distinguish it from nuclei of other cells

The pencil cell, like the ordinary epithelial cell, extends from the basement membrane to the lumen It possesses microvilli and terminal bars, and has a stratification of cytoplasmic organelles similar to that of the ordinary epithelial cell The basal wall of the pencil cells has papillary projections of cytoplasm which extend deep into the basement membrane The function of the pencil cell is unknown

The nucleus and cytoplasm of the basal cells are darker than those of the adjacent ordinary epithelial cells The plasma membrane of the cell is relatively straight The cytoplasm of these cells is more simply organized than is that of the other two cells previously described Small infrequent mitochondria and the endoplasmic reticulum are randomly distributed throughout the cytoplasm Only rarely are Golgi membranes or vesicles observed The function of the basal cell is not known

CONTRACTION OF THE GALLBLADDER AND COMMON BILE DUCT

The gallbladder shows spontaneous *rhythmical contractions* which occur at the rate of 2 to 6/min (in the dog), and also a *tonic contraction* which lasts from 5 to 30 min or more The rhythmical contractions (in the dog) are capable of producing a pressure change of 250 to 300 mm of water, which is about the maximal pressure at which bile can be secreted by the liver Rhythmical contractions of the common bile duct have been demonstrated also in animals

FUNCTIONS OF THE GALLBLADDER

Concentration and Storage of Bile—Secretion of Mucus. Gallbladder bile may have ten times more total solids than bile collected from the hepatic duct Water and inorganic salts are absorbed through the lymphatics and blood vessels of the gallbladder wall The composition of the absorbed fluid is virtually that of physiological saline Bile pigments, bile salts, and cholesterol are not absorbed to any appreciable degree under normal circumstances

It is undecided whether cholesterol is excreted by the normal gallbladder mucosa In

this connection, it may be mentioned that a pronounced diffuse deposition of a cholesterol ester in the connective tissue of the human gallbladder, ranging from 35 to 60% of its dry weight is seen as a pathological condition The tissue of the vesicle is stiff and greatly thickened as a result of its impregnation with lipid material The disturbances leading to this condition, which is spoken of as *cholesterosis of the gallbladder* or from the appearance of the mucosa, as the "strawberry gallbladder" are unknown The existence of this condition cannot, however, be used as evidence for the secretion of cholesterol by the gallbladder mucosa under physiological conditions

We owe the greater part of our knowledge of the absorptive powers of the gallbladder to the work of Rawdin and co-workers (1932 a and b) Rous and McMaster (1921), and more recently, that of Diamond (1962a, b, c) By means of a cannula placed in the bile duct, Rous and McMaster collected the bile as it came from the liver and compared its composition with the bile in the gallbladder The bilirubin percentages of the respective fluids were used as a measure of the degree of concentration that had been effected in the bile during its stay in the gallbladder The bladder bile was found to be darker, thicker and more "syrupy" than the bile collected from the ducts It contained from 3 1 to 10 8 times more bilirubin than the liver bile Absorption occurred with remarkable rapidity in some instances In one experiment about 50 ml of bile which entered the gallbladder was reduced to less than 5 ml in about 22 hr In experiments involving the drainage of the gallbladder through a cannula inserted into its fundus, it has been shown that the mere passage of bile through the organ causes a nearly fivefold concentration Inflammation of the gallbladder reduces or abolishes its concentrating power

Diamond (1962) has shown, in a series of beautifully designed experiments, that the gallbladders of fish and rabbit *in vitro* absorb sodium chloride and water in isotonic proportions The differences in composition between hepatic and gallbladder bile result secondarily from this transport process across the gallbladder mucosa Sodium and chloride are absorbed in 1:1 proportions and water movement is coupled to the salt movement. The differences in composition of hepatic bile and gallbladder bile of the dog are listed in Table 2 10

Table 2 10
Composition of canine hepatic and gallbladder bile

	Hepatic Bile	Gallbladder Bile
	mmoles	*mmoles*
Na	174	220-340
K	6 6	6-10
Cl	55 107	1-10
HCO_3	31 65	0 17
Bile acids	28 42	290 340
Ca	6	25-32
Mg	3 6	

After Diamond (1962a) From analyses by Rawdin, Johnston Riegel and Wright (1932) Reinhold and Wilson (1934) Wheeler and Ramos (1960)

Using the gallbladder of a freshwater fish, the roach, Diamond has studied its absorptive functions *in vitro* After the gallbladder is removed from the fish, the cystic duct is cannulated and the gallbladder is filled with a solution to be studied The cannula and gallbladder are then suspended in a beaker containing a bathing solution As a measure of absorption, the cannula and gallbladder are weighed at 5 to 10-min intervals Figure 2 40 shows the results of such an experiment when Ringer's solution was used to fill the gallbladder and also for the bathing solution in the beaker It was found that absorption occurred at a steady rate with the removal of approximately 38 mg/hr of the Ringer's solution in the gallbladder After 3½ hr, iodoacetate, a metabolic poison, was added to the solution in the beaker Absorption was promptly inhibited, as shown by no further loss of weight of the preparation, and so indicated that the absorptive process is metabolically dependent

Unlike most membranes that actively transport sodium or chloride, no potential difference is observed across the gallbladder mucosa during active absorption The "saline pump" is thus electrically neutral, due to the fact that for each sodium ion transported, one chloride ion is also transported. The limiting factor of the salt transport system appears to be the chloride concentration remaining in the gallbladder bile

It will be recalled that bile is in osmotic equilibrium with plasma The osmotic equilibrium exists even though the total electrolyte composition of bile may be as high as 600 mEq /liter This discrepancy in the total electrolyte composition and the osmolality of bile is observed in

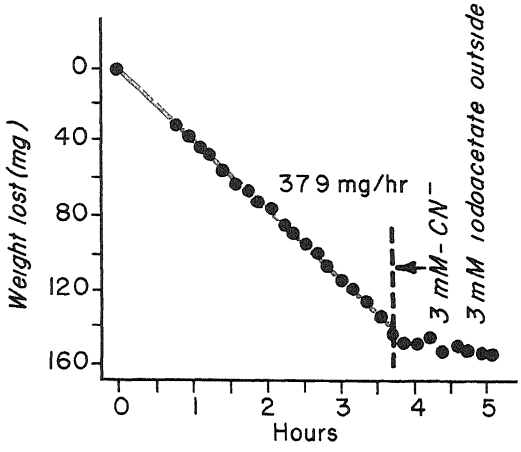


Fig 2 40 Absorption in the isolated gallbladder Both sides of the gallbladder were initially bathed in Ringer s solution Absorption from the gallbladder was determined by noting the decrease in weight of the preparation which averaged 37 9 mg /hr At 3 hr and 48 min NaCN and iodoacetic acid were added they blocked absorption as shown by no further decrease in weight (Redrawn from Diamond 1962a)

both common duct bile as well as in gallbladder bile, but it is much more marked in gallbladder bile that has been concentrated The explanation of this apparent discrepancy is found in the fact, previously mentioned, that bile salt exists as aggregates or micelles that are osmotically inactive As bile is concentrated, the formation of these macromolecular complexes (Juniper, 1964) reduced the osmotic effect of bile acid as well as the electrolytes in the bile, thus maintaining osmolality

The gallbladder mucosa also adds to the viscosity of the bile by the *secretion* of a thick mucinous material If the ducts are obstructed by ligation and the gallbladder is tied off, after some days, a clear, colorless fluid is found to have collected within the ducts This fluid, the so-called "white bile," is not uncommonly seen during an operation upon an obstructed bile duct associated with a functionless gallbladder The "white bile", under these circumstances, is furnished solely by the mucosa of the ducts It contains no pigment, bile salts, or cholesterol, and bears almost no resemblance to bile The secretion of the latter has been suppressed by the rise in pressure (to 300 mm or so of water) within the ligated duct system

If the gallbladder is healthy, and left in communication with the obstructed system, the sequence of events is entirely different Biliary stasis then causes thick, *greenish bile* to collect in the ducts and bladder as a result of the latter's concentrating activity, and of the mucinous material which it secretes After a lapse of weeks the imprisoned bile develops an almost tarry consistency The functions of the ducts and of the bladder are, therefore, antagonistic, the former tending to dilute, the latter to concentrate, the biliary fluid

Other Functions. *Other functions of the gallbladder*, subsidiary to its concentrating power, are the *reduction in the alkalinity* of the bile and the *equalization of pressure* within the biliary duct system Without the ability to absorb fluid and reduce the bulk of the bile, its power to equalize pressure would be negligible The amount of bile secreted in 24 hr is about 20 times or so greater than could be contained in the gallbladder The loss of its action in equalizing the pressure within the duct system is probably a factor leading to the dilation of the bile ducts, which so frequently follows removal of the gallbladder (cholecystectomy). After this operation, the flow of bile into the intestine is nearly continuous at first, but, later, the adaptation of the ducts permits intermittent discharge

The importance of the gallbladder in the control of pressure within the biliary ducts is apparent from the results of the experiments of Mann and Bollman (1925) They found that after ligation of the common duct in dogs, a rise in the bilirubin concentration of the blood did not occur until 24 to 36 hr had elapsed, and jaundice did not appear for 2 days If, on the other hand, the gallbladder was removed at the time that the duct was ligated, jaundice was fully developed within 24 hr due, presumably, to the rise in duct pressure and the "regurgitation" of bile into the blood

Menguy and associates (1958) have made measurements of biliary and pancreatic ductal pressures and the resistance of the sphincter of Oddi in dogs, under various conditions In intact dogs, the average biliary pressure was found to be 71 mm of water When pressures were simultaneously measured in the pancreatic ducts, the pancreatic intraductal pressure was always greater than the pressure in the common bile duct. After cholecystectomy, the pressure in the common bile duct was increased to an average of 111 mm. of water. This increased biliary pressure sometimes exceeded that measured in the pancreatic ducts. It was

also noted that the biliary pressure was much less stable after cholecystectomy The effects of a meal varied, according to whether or not the gallbladder was present In intact animals, a meal always caused a moderate rise in pressure due to contraction of the gallbladder In cholecystectomized animals, the already high fasting biliary pressure usually dropped after a meal The resistance of the sphincter of Oddi was markedly diminished after feeding

FILLING AND EVACUATION OF THE GALLBLADDER

The entrance of bile into the duodenum is blocked during fasting by the tonic contraction of the sphincter muscle (sphincter choledochus) at the duodenal end of the duct As the bile accumulates within the duct system, its pressure rises, and, reaching a height of 50 to 70 mm of water, forces its way along the cystic duct into the gallbladder During fasting, therefore, the viscus becomes gradually distended with retained bile

As the wall of the gallbladder is relatively thin, and its muscle fibers so sparse, it may seem unlikely that it could exert the pressure required to discharge its contents – especially since the gallbladder is evacuated with considerable difficulty by manual compression It has, however, been proved quite definitely, as a result of evidence derived from several modes of investigation, that the contractions of the gallbladder itself, despite the apparent muscular weakness of its walls, are responsible for the expulsion of its contents The times of emptying of the gallbladder are related to gastric digestion During fasting, it remains distended with bile.

When the walls of the gallbladder contract, bile is discharged along the cystic and common ducts into the duodenum. The sphincter (or sphincters) guarding the lower end of the common bile duct normally can withstand a pressure of 100 to 200 mm of water, but the pressure developed by the contractions of the gallbladder in dogs was shown by Mann and his associates to amount to over 250 mm of water Moreover, it is probable that relaxation of the sphincter occurs as part of a coordinated mechanism when the bladder wall contracts The duodenal muscle surrounding the oblique intramural portion of the common bile duct is capable, when contracted, of offering a resistance of over 750 mm of water. Since this is much higher than the pressure which contractions of the gallbladder can exert, the flow of bile is completely blocked during contractions of the duodenal muscle Therefore, during the evacuation of the gallbladder and during active duodenal movements, the bile may be observed to enter the duodenum in squirts This is not due to the "milking" action of the peristaltic movements of the bowel but is the result of the alternate blockage and release of the duct, the duodenal movements are incapable of causing any flow of bile when the gallbladder is not contracting

MECHANISM CONTROLLING EVACUATION OF THE GALLBLADDER

Nervous It has been mentioned that relaxation of the sphincter of Oddi probably occurs as the gallbladder contracts A coordinated action of this nature points to a nervous mechanism The latter may depend upon intrinsic nervous plexuses in the walls of the biliary passages The gallbladder contractions initiated by the stimulus of a meal might similarly be due to short reflexes through the intrinsic plexuses of the stomach or duodenum and biliary tract Nervous mechanisms are also indicated by the following observations Electrical stimulation of the stomach and duodenum in animals is followed by contractions of the gallbladder Contractions are occasionally induced by psychic influences, e g , the smell or taste of food The gallbladder also responds to experimental excitation of the vagus or sympathetic nerves Experiments attempting to demonstrate the precise actions of the extrinsic nerves upon the gallbladder movements have, however, given very conflicting evidence. Several observers have obtained weak motor effects from both vagal and sympathetic stimulation A motor action of the latter is also indicated by the fact that adrenaline is excitatory.

Hormonal Even in animals, nervous mechanisms are not essential to gallbladder activity This is evident from the fact that the reaction to the introduction of fat into the duodenum occurs after severing all nervous connections between the biliary and gastrointestinal tracts and between the former and the central nervous system That gallbladder contractions can occur under such circumstances suggests a hormonal or humoral mechanism Boyden (1928) found that the blood of an animal (taken at the height of digestion) injected into a

fasting animal causes the evacuation of the bile, blood from a starved animal has no such effect

Ivy obtained an acid extract from the musoca of the upper part of the intestine which caused contraction of the gallbladder when injected intravenously into animals Acid and other substances which are excitatory, when placed in the duodenum or when fed, act by causing the production or liberation of a hormone in, or from, the intestinal mucosa Ivy and Oldberg (1928) named this hormone "cholecystokinin "

The Actions of Drugs upon the Motility of the Gallbladder Epinephrine, pitressin, histamine and mecholyl stimulate the smooth muscle of the gallbladder, whereas morphine, ergotamine and atropine are inhibitory

Gallstones—Cholelithiasis

On the basis of autopsy data, it has been estimated that approximately 15% of Caucasians have gallstones The incidence of gallstones increases with age, cholelithiasis is rare in children and below the age of 20 years, but after the fourth decade, the incidence of gallstones is 25% and after the seventh decade, 50% Gallstones occur twice as commonly in women, and even in children who develop gallstones the ratio of females to males is 2 1 In children who have gallstones, the possible existence of hemolytic anemia should be considered strongly In both humans and animals, it has been observed that the incidence of gallstones in females is greatest in those who have borne offspring

The presence of gallstones is usually considered to be indicative of a diseased gallbladder The incidence of gallstones is higher in patients with diabetes, hemolytic anemia, and cirrhosis of the liver In patients with pancreatitis, approximately one-third will have cholelithiasis Except when producing obstruction of the biliary ducts, gallstones cause few symptoms Although the incidence of gallstones is extremely high (65 to 100% in various series reported) in patients with carcinoma of the gallbladder, only 1 to 2% of patients with gallstones develop carcinoma of the gallbladder

Gallstones are composed of constituents of the bile which have been thrown out of solution Cholesterol is present in greater or less amounts of the commoner varieties of gallstone Some stones may be composed almost entirely of this substance In other types of stone, bile pigment or calcium is an important or the predominant constituent Gallstones, therefore, differ considerably in size, color and inner structure, according to the materials of which they are composed The great majority of stones are formed in the gallbladder, but they also may form in the hepatic duct, or even in the smaller ducts within the liver

THE FORMATION OF GALLSTONES

The mechanism of gallstone formation is not clearly understood but the following are recognized as being the most important factors to be considered 1) *injury*, especially of an infective nature, to the gallbladder wall, 2) disturbance in *cholesterol metabolism*, 3) *stasis* of the bile, and 4) *reaction* of the bile

EXPERIMENTAL PRODUCTION OF GALLSTONES

Little experimental work has been done on the pathogenesis of cholelithiasis As in man, gallstones are not infrequently found in animals which possess gallbladders Small animals, which have been used as experimental models for inducing cholelithiasis in the laboratory, include the guinea pig (Okey, 1942), the hamster (Dam and Christensen, 1952), and the mouse (Tepperman et al , 1964) To induce cholelithiasis experimentally, manipulation of the diet of the animal has been the method most commonly employed For example, Okey found that with a high cholesterol diet with added vitamin B, gallstones would develop in 80% of guinea pigs A high cholesterol diet in itself would not induce cholelithiasis In both hamsters and mice, the ratio of cholic acid to cholesterol in the diet, and in the secreted bile, appears to be a critical factor It has already been mentioned that in normal bile, the cholesterol bile acid ratio is between 1 20 and 1 30 When this ratio falls to 1.13, deposition of cholesterol occurs In mice, a high incidence of cholelithiasis can be induced by simply adding 1% cholesterol and 0 5% cholic acid to the animal's regular laboratory chow diet On such a diet, the mice will begin to show stones within 3 to 4 weeks Females tend to develop stones earlier than males Tepperman and his associates (1964) have pointed out that the rate of bile flow is also important in the production of cholelithiasis in mice They found that female mice had a significantly higher rate of bile

production than did males, and that estrogen increased the rate of bile flow in the male while androgen decreased the rate in the female. This recalled the well-known relationship between pregnancy and gallstones.

In both hamsters and mice which have developed cholelithiasis due to the addition of cholesterol and cholic acid in the diet, the stones tend to disappear after several months on a normal diet (Caldwell et al., 1964). The spontaneous disappearance of gallstones in man has also been reported (Miller, 1956; Dworken, 1960). The factors involved in the dissolution of gallstones are not understood. The explanation for the disappearance of a few small stones might be that they are passed through the biliary system into the intestine. This is not the explanation for large stones, however. Although apparently very rare, gallstones, on occasion, can spontaneously disappear.

Further study of the various animal models for the production of cholelithiasis should yield information that will be helpful, not only in the understanding the etiology of cholelithiasis in humans, but also in prevention and therapy after gallstones have developed.

8 / SECRETION AND ABSORPTION IN THE INTESTINE

SECRETION

The food substances in the gastric chyme are far from being ready for absorption The products of peptic digestion of protein are still relatively large molecules containing many amino acids Starch digestion is incomplete and fat digestion has barely begun In the intestine, the enzymes of pancreatic juice carry the digestion of protein a step further but leave most of it still in the form of peptides, which need further digestion before being absorbed Starch and dextrins are reduced to maltose and some glucose Presumably, fat digestion is accomplished by pancreatic lipase but even this may be aided by lipase from the intestine Final food preparation for absorption is accomplished by enzymes found within the lumen of the small bowel and within the epithelial cells of the mucosa lining the small bowel

Secretions of the Small Intestine

The structure of the intestinal mucosa is particularly adapted to provide an extremely large surface area to the intestinal contents and thus enhance the process of absorption This increase in surface area is accomplished by the *villi* of the mucosa, and the brush border (composed of *microvilli*) of the epithelial cells covering the villi At the base of the villi are small tubular glands known as the *crypts of Lieberkuhn* It has long been taught that the "intestinal glands" secreted and that the villi absorbed Many enzymes have been identified in the contents of the small bowel and it was assumed that they were the products of secretion by the intestinal glands This secretion, known as the *succus entericus*, was supposed to complete the digestion of carbohydrates, proteins, and fats, in order that the products of digestion could be absorbed by the villi Recent evidence has cast considerable doubt on this concept of intestinal secretion and absorption It now appears that the primary function of the "intestinal glands" is to provide for the constant replacement of epithelial cells for tips of the villi In addition, it is now known that carbohydrates, proteins, and fats do not have to be completely reduced, respectively, to monosaccharides, amino acids, and fatty acids, prior to absorption

THE INTESTINAL GLANDS

The mucous membrane of the small intestine is covered everywhere with minute projections (about 1 mm in height and somewhat less in diameter) called intestinal villi The villi are invested with a layer of columnar cells, of the type characteristic of the intestinal mucosa set upon a basement membrane, beneath which is a fine layer of a smooth muscle fibers continuous with the muscularis mucosae In the center of each villus is a lymphatic vessel continuous with the lymph vessels of the mucous membrane This lymphatic is somewhat enlarged into a small sinus lined with endothelium In some of the larger villi there are several lymphatics Between the epithelium and the central sinus is a network of blood vessels Contraction of the smooth muscle of the villus causes a characteristic type of movement Between the villi are the openings of the *intestinal glands*, or *crypts of Lieberkuhn*, which are simple tubular glands The crypts do not penetrate the muscularis mucosae The epithelium covering the villi is of the same type as that lining the crypts, and consists of a single layer of columnar cells with certain specialized cells interspersed among them The free end of each columnar cell, next to the lumen of the intestine, is provided with a specialized cuticular border, resembling the brush border of certain renal tubular cells The brush border of the surface epithelial cells of the villi is composed of *microvilli* approximately 1 μ in length and about 0 1 μ in width Zetterqvist (1956) has estimated that each columnar absorbing cell in the mouse jejunum possesses about 600 microvilli and that 1 mm ^{2} of absorbing surface will contain 50,000,000 such structures

Here and there among the columnar cells are goblet cells that secrete mucus and specialized cells that stain with silver and are called *argentaffin* or enterochromaffin cells. Other cells with a large acidophile nucleus known as *Paneth* cells are also present. The argentaffin cells are of particular interest because they secrete serotonin, or at least synthesize its precursor 5 hydroxytryptophane (Erspamer and Asero 1952). The functions of serotonin in the body are not known, but it has many powerful pharmacological actions, among which are stimulation of gastrointestinal motility and inhibition of gastric secretion (Olson and Gray, 1958).

In the bottoms of the crypts many of the epithelial cells may be seen to be undergoing mitosis, and at the tips of the villi one can see that cells are being shed into the lumen of the intestine (Leblond and Stevens, 1948). Thus, there is going on in the intestinal mucous membrane a process of continuous replacement of the epithelial lining: new cells being produced in the bottoms of the crypts, and the older cells being continuously shed at the tips of the villi. The rate of renewal of shed epithelial cells has been estimated by means of autoradiography using tritiated thymidine to label the deoxyribonucleic acid (DNA) of the nuclei (Leblond and Walker, 1956). When tritiated thymidine is injected into an animal, it is rapidly incorporated in the cells of the crypts undergoing mitoses. The migration of the labeled cells up the villus can be observed by obtaining biopsies at various times after the injection, and exposing thin sections on a photographic emulsion. In man, the time required for this migration has been estimated as 3 days. Thus, every 3 days the lining of the small bowel is replaced due to the rapid turnover of epithelial cells. Crosby (1961) has estimated, on the basis of calculations of the epithelial cell mass of the small bowel of man, that about 250 g of epithelial cells are shed into the lumen of the small bowel each day.

In the first part of the duodenum, in addition to the regular intestinal glands, are special mucous glands known as *Brunner's glands*. These are similar in structure to the pyloric glands of the stomach. They are made up of long tubules, which branch frequently and are often tortuous, and which penetrate the muscularis mucosae. Their ducts empty into the crypts of Lieberkühn. Brunner's glands are very numerous in the duodenum between the pylorus and the entrance of the bile and pancreatic ducts. Below this level they are seen less frequently and none are found beyond the duodenojejunal junction.

INTESTINAL SECRETIONS OR SUCCUS ENTERICUS

The composition of the fluid secreted by the intestinal mucous membrane varies somewhat in different parts of the intestine. For example, in the duodenal area, due to the secretion of the mucous glands of Brunner, the secretion contains more mucus than is found elsewhere. In general, the secretion of the small intestine is a thin, colorless, or slightly straw-colored fluid, somewhat opalescent and containing flecks of mucus. On centrifuging and examining the sediment, it can be seen that the cloudy appearance of the juice is partly due to mucus and partly due to cellular debris, including some intact cells of the type characteristic of the intestinal mucous membrane (Florey et al., 1941). When collected from the duodenum only, the juice has a more glairy appearance and a heavier consistency, and evidently contains a greater proportion of mucus. The composition of the juice consists mainly of water, inorganic salts, and organic material.

The fasting intestinal fluid is the resultant of fluxes into and out of the intestine, and what is collected and measured is a *net* resultant.

The concentration of sodium, potassium, calcium, and total anions is relatively constant, and in both jejunum and ileum resembles the serum. Cl and HCO_3 display a reciprocal relation, in jejunal juice HCO_3 is low, and is high in the ileum.

The reasons for the differences in HCO_3, Cl are not established. It has been suggested that the difference between jejunum and ileum is due to a diminishing ability to secrete hydrogen ions along the gastrointestinal tract. Recently Fordtran and his colleagues (1968) have suggested a double exchange mechanism, Na^+ for H, and chloride for bicarbonate for electrolyte movement across the ileum, based on a nonelectrogenic transport system. It has also been argued that passive diffusion could explain the electrolyte movements.

Except in the duodenum, the quantity of fluid secreted by the small intestine is never very great, a few milliliters per hour being the usual amount that it is possible to obtain under conditions of stimulation. In the dog, pig, rab-

bit, and goat, spontaneous secretion of the Brunner's glands, amounting to 0 5 to 2 0 ml /hr , has been demonstrated Spontaneous secretion apparently does not occur in the cat In dogs, where more extensive studies have been done, the response of the Brunner's glands to feeding results in only 3 to 4 ml /hr It is difficult to determine with accuracy the amount actually secreted because of the tremendous absorptive capacity of the intestine Many times more than the amount that it is possible to collect may be secreted and reabsorbed during an observation period It is evident that the mucosa is capable, under some circumstances, of passing enormous amounts of fluid into the lumen of the intestine as shown by the great water loss that occurs through the intestine in pathological states such as cholera, diarrhea, or intestinal obstruction

The organic matter of the juice consists of mucus, enzymes, and cellular debris The source of the enzymes is unknown but it is assumed that they are produced by the columnar cells, although the argentaffin cells or the Paneth cells may be responsible for some of them There is no evidence that the mucosa of the small bowel *secretes* enzymes in the sense that secretion occurs in the gastric mucosa or in the pancreas The exfoliation of epithelial cells into the bowel lumen, and their subsequent dissolution probably accounts for the presence of enzymes in the intraluminal contents of the small intestine, although Florey and his colleagues have concluded that only enterokinase, the enzyme which converts trypsinogen to trypsin, and amylase were secreted into the lumen of the gut A really formidable array of enzymes have been reported as occurring in intestinal secretion A pepsin-like protease (from the duodenum only), an amylase, a lipase, at least two peptidases, sucrase (invertase), maltase, lactase, enterokinase, alkaline phosphatase, nucleophosphatases, and nucleosidases have been described Immunoglobulins have been identified in the secretions of the intestine Immunoglobulin A (IgA) is the predominant form throughout the intestinal tract of man, and is believed to be transferred from the blood to gut lumen by the addition of a "secretory piece" or transport "T" piece

The studies of Wright, Jennings, Florey and Lium (1940) raised a question as to the source of the enzymes ordinarily found in intestinal secretion They reported that when they took precautions to prevent injury to the mucous membrane while collecting the secretion, and then immediately centrifuged the collected secretion so as to remove cellular debris, only amylase and enterokinase were consistently present in the supernatant fluid They consider that those enzymes, and possibly maltase, are secreted by the intestinal glands in the usual sense The others, they believe, are intracellular enzymes which exist in the cells of the mucous membrane and appear in the juice only as a result of the shedding of these cells and their disintegration in the intestine

CONTROL OF BRUNNER S GLANDS

Secretion of Brunner's glands is controlled by nervous and humoral mechanisms which affect the other intestinal glands only to a slight extent Their secretion increases on stimulation of the vagus nerves or the administration of parasympathomimetic drugs such as pilocarpine and physostigmine, it is not augmented by stimulation of the sympathetic nerves (Florey and Harding, 1934) The amount of secretion is increased after meals (Blickenstaff et al , 1949), even in transplanted denervated pouches of the duodenum (Florey and Harding, 1935b) This proves that there is a humoral mechanism for regulating the secretion, in addition to whatever nervous mechanisms there may be Florey and Harding (1935a) were convinced that it was secretin More recent studies indicate that in addition to secretin, gastrin, pancreozymin, and glucagon stimulate Brunner's glands as well

Florey and Harding (1933) emphasized the protective function of the secretion of Brunner's glands and considered it an important factor in preventing duodenal ulceration Since, like other intestinal secretions, it contains enterokinase it helps to activate the trypsinogen of pancreatic juice

CONTROL OF SECRETION OF SUCCUS ENTERICUS

It has long been known that local chemical and mechanical stimuli can evoke secretion by the small bowel The neural and humoral mechanisms regulating secretion by the small bowel are poorly understood Except for "paralytic secretion," discussed below, there is little evidence that neural and humoral mechanisms normally play a significant role

Nervous Control Wright and his coworkers (1940) were able to obtain an increased secretion from the duodenum on stimulation of the vagus in decerebrated or decapitated animals, but were unable to obtain clear evidence that vagal stimulation caused secretion from the jejunum or ileum

Stimulation of the sympathetic causes no secretion, but cutting the nerves results in a marked increase in secretion This *paralytic secretion*, as it is called is increased by physostigmine and inhibited by atropine, it is therefore considered to be dependent upon some cholinergic mechanism in the intestine In general, parasympathomimetic drugs increase the paralytic secretion, and sympathotaimetic drugs tend to inhibit it From these observations we can conclude that secretion is augmented by a cholinergic mechanism One interpretation of the paralytic secretion is that section of the sympathetic nerves results in either unopposed acetylcholine stimulation or dilation of the blood vessels of the intestine, which increases the blood flow An increase in blood flow could provide a greater amount of fluid for secretion and, presumably, produce augmentation of secretion on section of the sympathetic nerves (Starling, 1906) Babkin (1950) has pointed out that cutting the sympathetic nerves to the small intestine increases its motility, which undoubtedly has a massaging effect on the mucous membrane

Humoral Control. The most effective stimulus for secretion of succus entericus is local, mechanical, or chemical stimulation of the intestinal mucous membrane Such stimuli are always present in the digesting intestine, due to the presence of chyme and the food particles which it contains Feeding is not very effective in increasing the rate of intestinal secretion from an isolated intestinal loop, provided the nerve supply is intact, but if the sympathetic nerves have been severed, a noticeable increase in secretion takes place when the animal is fed Nasset (1938) and his coworkers have prepared an extract of the intestinal mucous membrane which is free from secretin, from vasodilator material, and from toxic substances, but which causes an increase in the secretion of a denervated jejunal loop when given intravenously The material is thought to contain a hormone which augments the intestinal secretion, it is called "*enterocrinin*," but its existence has not been generally accepted by workers in the field

THE FUNCTION OF THE SUCCUS ENTERICUS

Enzymatic digestion of food, in all its phases, is a hydrolytic process, therefore, an abundance of water to serve as one of the reacting substances is essential The succus entericus provides this water in the area in which the major part of digestion of food occurs Water is also necessary to serve as a solvent and as a medium of suspension and transport for the solids which are dissolved or suspended in the chyme The enzymes, either in the secretion or in the cells, complete the digestion of protein and carbohydrate by reducing the peptides which result from peptic and tryptic digestion to amino acids, and the disaccharides which result from amylolytic digestion of starch to glucose Other disaccharides that may be present in the food are likewise digested, liberating their constituent monosaccharides In fat digestion, the succus entericus serves as a source of water and as a medium for suspension and emulsification of the fat particles, it may provide a certain amount of lipase

Secretion of the Colon

The mucosa of the mammalian cecum and colon is in many respects similar to that of the small intestine, crypts are present but there are no villi As in the small intestine, the cells at the bottoms of the crypts exhibit numerous mitoses, suggesting the continuous replacement of shed epithelial cells Goblet cells are more numerous than in the small intestine, and the epithelial cells between the goblet cells differ among themselves, but in the majority of cases, the protoplasm is clear and free from secretory granules The secretion is scanty (Wright et al, 1938), as is to be expected from the fact that a major function of the large intestine is absorption of water The decrease in volume of intestinal contents that occurs in this area indicates that secretion does not keep pace with absorption Actually, it is usually not possible to collect any secretion from a colonic fistula unless the mucous membrane is stimulated mechanically by insertion of a catheter or other instrument Even with this stimulation only a few tenths of a cubic centimeter per hour can be collected When anything can be collected, it consists of watery fluid with clumps of white mucus; often it is viscous and opalescent The reaction is alkaline, due to sodium bicarbonate which may be present in concentrations as

high as 80 to 90 mEq /liter, according to DeBeer et al (1935)

NERVOUS CONTROL OF COLONIC SECRETION

Florey (1930) and Wright et al (1938) found that stimulation of the nervus erigens caused the colon to secrete a clear mucoid fluid In contrast to absence of spontaneous secretion, stimulation produced as much as 55 ml in one 8-hr period from the distal half of the colon, the average rate was about 5 ml /hr Reflex secretion was observed on stimulation of the cut end of one nervus erigens, the other being intact, the reflex center was located in the lumbar spinal cord It should be recalled that the nervus erigens supplies only the distal portion of the colon with parasympathetic fibers, the parasympathetic supply for the proximal colon comes from the vagi Acetylcholine and pilocarpine increase colonic secretion, atropine inhibits it Cutting the sympathetics does not produce "paralytic secretion" in the large intestine, but stimulation of the sympathetics diminishes the secretory response to stimulation of the nervus erigens

ABSORPTION

The mucous lining of the small intestine is one of the most remarkable organs of the body It secretes a variety of hormones which help to regulate gastric secretion and motility, the secretion of the pancreas and of the intestine itself, and the movements of the gallbladder and the intestinal villi During absorption, the mucosa cells perform a variety of hydrolytic and synthetic operations upon the food substances passing through them, greatly modifying the character of the absorbed material Even while producing their own secretions, the epithelial cells are transferring water and dissolved substances from the lumen of the intestine into the interstitial fluid to be taken up by the blood This activity is highly selective and frequently involves the expenditure of considerable energy Probably nowhere, except possibly in the liver, is so great a variety of functions performed by a single tissue

The ultimate aim of digestion is absorption Only after the products of digestion have passed through the epithelial lining of the digestive tract and entered the blood stream are they able to serve their purpose of providing nourishment for the tissue cells of the body Absorption of such materials takes place almost entirely from the small intestine Although the stomach is capable of absorbing small amounts of water and certain food substances, notably glucose, and some foreign substances such as alcohol, the absorption that occurs there is negligible Although absorption of nutrients occurs mainly in the small intestine, water and salts are absorbed in significant amounts from the large intestine

The mucous lining of the small intestine is admirably adapted to its absorptive function The structure of the epithelium has been described in connection with the discussion of intestinal secretion It has been estimated that the presence of the villi results in an increase of 7- or 8-fold in the surface area of the mucous membrane of the intestine The microvilli increase the surface area by another 20-fold (Fig 2 41)

Methods for Study of Intestinal Absorption

In acute experiments, one places the material, the absorption of which is to be studied, in an isolated loop of intestine and, after a measured time, removes it and observes the volumetric or chemical changes that have taken place The Thiry or Thiry-Vella loop may be used in the same manner In the case of substances that can be readily detected in the blood, the material can be placed in the intact

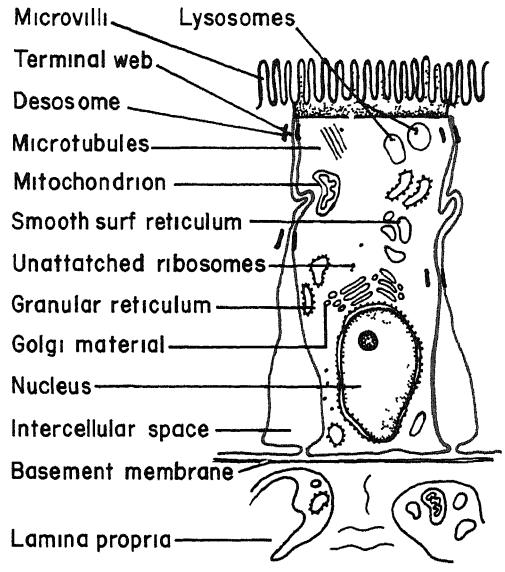


Fig 2 41 Schematic diagram of the intestinal absorptive cell (From Trier and Rubin 1965)

intestine and its absorption estimated by the rate of its appearance in the circulating blood, in the portal blood, or in the intestinal lymph, if the experiment is of such a nature as to make these vessels available

In man, intestinal absorption can be studied by an intubation technique A multilumen tube, or two or more separate small polyethylene tubes may be used After localization in the desired segment of intestine, the material to be studied, along with a nonabsorbable reference "marker," is introduced through the lumen of the proximal tube at a constant rate Sampling is done through one or more of the distal tubes, whose openings are a measured distance from the opening of the proximal tube used for introduction The concentration of absorbable material recovered is compared to the concentration introduced, and a correction is made for any change in volume as indicated by the change in concentration of the "marker" Polyethylene glycol (PEG) and phenol red are commonly used as "markers" for such studies

The development of in vitro techniques for the study of excised segments of intestine has greatly increased our knowledge in recent years A variety of techniques for this purpose have been described, all depend upon maintaining the vitality of a portion of the intestine that has been removed from the body by keeping it in contact with oxygenated fluid of appropriate composition The preparation can be arranged so that the fluid inside the intestine can be kept separate from that on the outside and the changes in composition of the internal and external fluid are used to measure the degree of absorption by the mucosa A small segment of intestine can also be turned inside out, as described by Wilson and Wiseman (1954), and is known as the everted sac method The method has the advantage of collecting high concentrations of the absorbed material in the small volume contained inside the serosal lined sac

Absorptive Functions of the Intestines

ABSORPTION OF WATER

The human intestine absorbed about 8 liters of water, but until recently it has been impossible to distinguish between the water that was secreted into the intestine and water that left the intestine by absorption. Changes observed in the volume of water in the intestine do not represent total absorption of water but merely the difference between the volume of water secreted into the intestine and that absorbed Since there is reason to believe that water moves freely through the intestinal mucosa, it is essential that we know how much water entered the intestine in a given time if we are to make an intelligent estimate of the amount leaving by way of absorption in the same time It has been possible to get around this difficulty, and to determine accurately the amount of water leaving the intestine by absorption in a unit of time, by the use of heavy water Water containing heavy hydrogen can be readily detected in the blood, also, the amount remaining in the intestine can be accurately measured Since, in a short period of time the water secreted into the intestine will be normal water, the change in heavy water content represents water actually absorbed Or if, as in experiments on human subjects, it is not possible to determine the amount of water remaining in the intestine, the amount of heavy water appearing in the blood can be used as an estimate of the water absorbed

Scholer and Code (1954) studied absorption of water from the human stomach and small intestine using heavy water as a tracer They found that when 50 g. of water were given, 2 5% of this amount was absorbed per minute, so long as it remained in the stomach When the water was placed in the small intestine, it was absorbed at about 10 times that rate, 26 1% of the dose being absorbed per minute If the water remained in the stomach, it required 34 2 min to absorb 67% of the administered amount; whereas in the intestine, this percentage was absorbed in 3 7 min In the intestine, 95% of the administered water was absorbed in 10 min In these experiments only absorption was measured, and the net volume change was not observed, hence, no account was taken of the water that might have been secreted into the stomach or intestine during this same period of time One interesting fact brought out in these studies is the extremely rapid absorption of water from the intestine Another is that there is an appreciable absorption of water from the stomach. Considering net absorption only, that is, the decrease in the volume of intestinal contents, it appears that water is absorbed more rapidly from the terminal ileum than from other parts of the small

intestine Most of what remains is absorbed from the large intestine, except the rectum

The movement of water is generally held to be passive, determined by osmotic and hydrostatic pressure gradients, which in turn are the result of the permeability characteristics of the intestinal membrane, active transport of solute, and tissue pressures due to the effects of active transport and blood and lymph pressures Absorption is accelerated by an increase in hydrostatic pressure up to the point where there is interference with blood and lymph flow

In general terms, modern studies treat the intestinal membrane as a lipid membrane perforated with water-filled pores, or a series of membranes In the model of Curran and Schultz (1966), the active transport of sodium leads to the absorption of water since sodium is concentrated in a continuing compartment, which thus establishes an osmotic gradient Water follows this osmotic gradient and isotonic flow ensures

MONOVALENT ELECTROLYTES

In considering the absorption of water and salt, the porosity of the intestinal membrane would obviously influence these movements By determining the reflection coefficient (the ratio of observed to theoretic osmotic pressure) for nonlipid solutes of varying molecular sizes, Fordtran and his colleages (1968) have made estimates of the "pore" size, and from a study of the filtration coefficient (the flow of water induced by an osmotic gradient) suggested that the "pores" in the jejunum were either more numerous or larger than those in the ileum

Sodium. Sodium is considered to be actively transported, i e , absorbed from lumen to blood or intestinal fluid against an electrochemical gradient The transmucosal potential difference of 4 to 8 mV (serosa positive to mucosa) has been attributed to this active sodium transport Further sodium absorption is accelerated by actively transported nonelectrolytes such as glucose and some amino acids such as alanine

Chloride Chloride is believed to follow sodium absorption passively but the situation may be more complex In rat jejunum, chloride is absorbed more slowly than sodium, in ileum at least as fast In man, with perfusion studies, Fordtran et al (1968) found in the jejunum that sodium was absorbed against a small concentration gradient and influenced by glucose In the ileum, sodium was absorbed against a sharp gradient, unaffected by glucose In the upper small intestine, they concluded that the majority of sodium absorption was accomplished by bulk flow along osmotic gradients

Potassium. Potassium is believed to move passively in response to concentration gradients

Bicarbonate Bicarbonate has been alluded to above, being lower than plasma in jejunal contents and high in ileal, and its reciprocal relationship to chloride has been commented upon The absorption of HCO_3^- could be secondary to the secretion of H^+ ion, formation of CO_2 and its diffusion from the lumen, or perhaps to the double exchange model (Na^+ for H^+, and Cl^- for HCO_3^-) mentioned before

ABSORPTION OF DIVALENT CATIONS

Calcium Calcium is absorbed in most species in the proximal small bowel by active process which involves two steps from the lumen into the mucosa, and from the mucosa to the underlying tissues Both steps require vitamin D, and the latter is dependent on oxidation phosphorylation At present, it is believed that vitamin D is converted into its active metabolite 25-hydroxycholecalciferol by the liver, and that this steroid induces the synthesis of a calcium-binding protein in the duodenal mucosa Parathormone plays a facilitory role in calcium absorption Calcium is absorbed more readily by individuals suffering from a calcium deficiency, indicating some control of calcium absorption by systemic factors Excess of magnesium interferes with calcium absorption, as does phytic acid (which occurs in cereal grains, and which may be responsible for the tendency of these grains to produce rickets when fed exclusively or in excess)

Magnesium. In man, magnesium is probably absorbed along the course of the entire small bowel Although calcium and magnesium are often thought to share a common transport system, workers have been unable to demonstrate magnesium transport against a concentration gradient using in vitro preparation of rat intestine

Iron. Like calcium, but unlike almost all nutrients, the absorption of iron is influenced by the body's needs, and it is held at present that the regulator of absorption is the columnar epithelial cell of the upper small bowel

In Crosby's formulation (1961) the columnar cell contains a variable amount of iron supplied by the body stores This deposit regulates within limits, the quantity of iron which can enter the intestinal cell Part of this accepted iron may proceed from the cell to body stores by way of the iron binding β globulin of the plasma (transferrin) and part fixed in the cellular cytoplasm is sloughed at the end of the cell's life span Within the cell there is an interplay of iron stored as ferritin iron-containing enzymes, and hemoglobin iron in two mucosal pools divalent iron for transport and trivalent ferritin iron which is concerned with temporary storage

It should be recalled that the bulk of dietary iron is derived from hemoglobin and organically complexed iron of food which must be digested to be available for absorption

ABSORPTION OF CARBOHYDRATES

Under the influence of salivary, pancreatic, and intestinal amylase, the carbohydrates of our diet are presented to the upper small bowel for absorption mainly in the form of various disaccharides and the monosaccharides glucose, fructose, and galactose Glucose occurs as the final product of starch digestion, and as one of the two products of hydrolysis of sucrose and lactose Fructose occurs as a product of sucrose digestion, and also in certain natural products, such as some fruits and in honey Galactose is one of the constituents of lactose Other simple sugars such as mannose and pentoses are quantitatively unimportant The three simple sugars, galactose, glucose, and fructose, are all actively absorbed from the small intestine, galactose more rapidly, and fructose less rapidly than glucose In addition, some disaccharides are also presented for absorption These include maltose, sucrose, and lactose derived from partially digested starch and ingested cane sugar and milk These, as will be pointed out below, are hydrolyzed to their corresponding monosaccharide by enzymes contained in the brush borders of the intestinal absorptive cell, which in general, is impermeable to polysaccharides

Absorption of Monosaccharides. The ability of the intestine to absorb glucose against a concentration gradient has been repeatedly demonstrated, i e , glucose can be absorbed when the concentration of glucose in the intestinal contents is lower than the concentration in the blood

It is interesting to note that although the pentoses are more diffusible than the hexoses, the pentoses are absorbed at a much slower rate proving again that the hexoses are actively absorbed With most sugars, particularly those that are not actively absorbed, the rate of absorption increases with increasing concentration in the intestinal contents With glucose and galactose however this is true only up to about a 5% concentration, which for these sugars is approximately isotonic with the blood Further increase in concentration of these sugars fails to increase or decrease their rate of absorption Furthermore if glucose and galactose are given simultaneously each suppresses the absorption of the other (Cori, 1931) so that the total amount of sugar absorbed is not more than would be absorbed if only one of the sugars were present These facts indicate that the mechanisms for active absorption can be saturated in a manner somewhat analogous to the tubular maximum for glucose in the renal tubules

Glucose is absorbed much more rapidly from the upper than from the lower small intestine Fisher and Parsons (1949) have shown that, in the rat, there is a linear gradient in the ability of the intestine to absorb glucose from the intestinal contents The rate of absorption is directly proportional to the distance of the segment in question from the ileocecal valve (Fig 2 42)

Using various segments of hamster intestine,

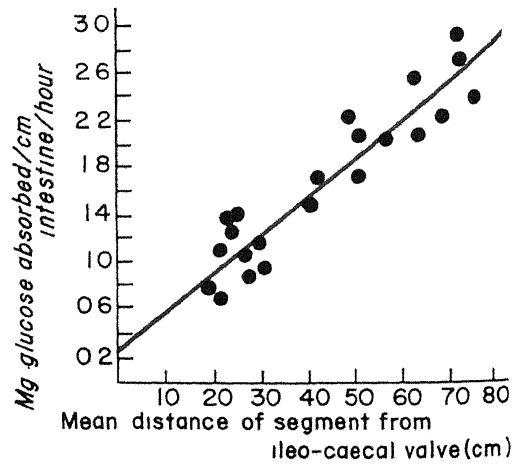


Fig 2 42 Glucose absorption in surviving intestine Relation between glucose disappearance from inner fluid and mean distance of segment from ileocecal valve in 1 hr experiments (From Fisher and Parsons 1949 Courtesy of Cambridge University Press)

Crane and Mandelstam (1960) have shown that the midportion of the small intestine is the site for the most active rate of absorption of monosaccharides as glucose Borgstrom and his associates in Sweden (1957) have studied the site of absorption of glucose and lactose in the small bowel of man They used an intubation technique, with polyethyleneglycol as a nonabsorbable indicator substance The subjects were fed a test meal containing 55 g of glucose and 20 g of lactose Samples were then collected from various levels of the small bowel, and it was found that absorption began in the duodenum and was completed by the time the meal had reached a level about 100 cm from the duodenum

As to the specific site of absorption in the mucosa, Kinter (1961), using autoradiographic techniques, has shown that absorption of glucose occurs mostly in the upper third of the villus Crane (1962) has presented evidence that the active transport of glucose actually occurs across the plasma membrane of the microvilli in the brush border

Absorption of sugars is diminished in the presence of sodium chloride deficiency such as may occur in adrenal cortical insufficiency, is accelerated in hyperthyroidism, and is depressed in hypothyroidism Pantothenic acid, thyamine, and pyridoxine, members of the B vitamin group, are essential for normal sugar absorption

For optimal absorption of glucose it is necessary that both sodium and potassium be present If sodium is completely replaced by potassium, lithium, magnesium, or mannitol, active glucose transport ceases If sodium transport is inhibited by cardiac glycosides, active glucose transport is also interfered with It is now thought that mechanisms of glucose and sodium transport are closely coupled (Crane, Miller, and Bihler, 1961) and dependent on one another Whether the sodium "pump" and the glucose mechanism are the same for entry into the cell is not known, however, Bihler, Hawkins, and Crane (1962) have suggested that two processes, both dependent on sodium, are involved in sugar transport *1*) an *entry* step that is sodium dependent but requires no energy, and 2) an *accumulation* step, also sodium dependent, but which requires energy to maintain a concentration gradient within the cell (Fig 2 43)

In addition to the specific dependence of glucose absorption and transport on the Na^+ ion, it is now known that there are specific structural requirements of the monossaccharide molecule for it to be actively transported The absorption of a great many naturally occurring and synthetic monosaccharides has now been studied in relation to the structural configuration of the molecule To be actively transported, the following specific structural requirements of the monosaccharide molecule are known *1*) The hydroxyl (OH) group at carbon-2 must be present and have the same stereoconfiguration as exists in D-glucose If the hydroxyl group at carbon-2 is replaced by H (as in 2-deoxyglucose), or if the H of the OH group is replaced by a methyl group (as in 2-O-methyl glucose), active transport does not occur 2) A pyranose (6-membered) ring must be present, as in glucose and galactose *3*) One or more carbons must be attached to carbon-5 The structural requirements necessary for the active transport of monosaccharides are illustrated in Figure 2 44

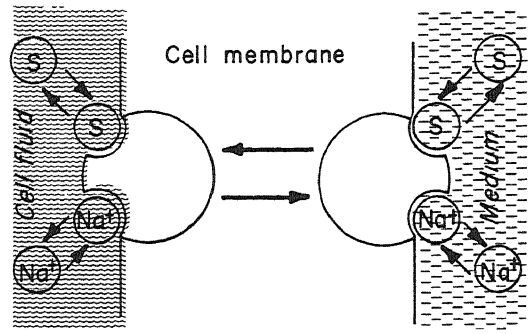


Fig 2 43 Model of mobile carrier with two sites one specific for substrates and one specific for Na (From Crane 1965)

From the above, it might be reasoned that 5-carbon sugars such as xylose will not be actively transported, yet recent data suggests that *d*-xylose is actively absorbed If a monosaccharide is a 7-carbon sugar, it may be actively transported if it has the specific structural requirements outlined above The hydroxyl group at carbon-1,3,4, or 6 is not essential for active transport Thus, phosphorylation which occurs at carbons-1 and -6 is not essential for active transport In other words, monosaccharides with the necessary specific structural requirements can be actively transported in the absence of phosphorylation Fructose is absorbed as such by passive diffusion and converted in part in the cell to glucose

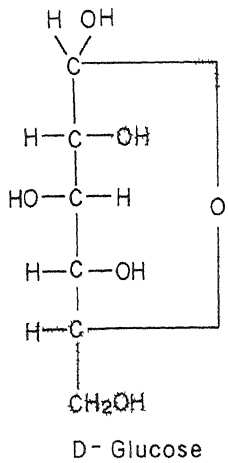


Fig 2 44 The minimal structural requirements for the active transport of monosaccharides by the intestine (From Isselbacher and Senior 1964)

Absorption of Disaccharides It has long been believed that disaccharides were hydrolyzed in the lumen of the small bowel by disaccharidases secreted by the small bowel into the "succus entericus" There is now much evidence that this is not the case, and that some disaccharides are absorbed and hydrolyzed *within* the epithelial cell by specific intracellular enzymes Much of the evidence in support of intracellular hydrolysis of disaccharides has been due to the work of Miller and Crane (1961) These investigators incubated inverted segments of hamster small bowel in a sucrose solution, and found that glucose accumulated in the epithelium to a concentration 20 to 30 times that in incubating medium This suggested that the sucrose was hydrolyzed after absorption, rather than before In another experiment, maltose was used in the incubation medium but glucose oxidase was added also A molecule of maltose results in two molecules of glucose on hydrolysis, but in the presence of glucose oxidase, the glucose would be oxidized and unavailable for absorption if hydrolysis occurred prior to absorption Their findings were the same as when sucrose was used in the incubation medium and glucose accumulated in the tissue, in spite of the fact that glucose oxidase was in the medium Other experimental evidence provided by Miller and Crane, in support of intracellular hydrolysis of disaccharides, is their demonstration that practically all of the maltase and invertase activity of the small bowel mucosa can be found in the isolated brush border of the epithelial cells.

Within the past few years a new clinical entity has been described in patients who are unable to absorb certain disaccharides The condition is now recognized as an *hereditary disaccharide intolerance* (Dahlqvist 1962) and is due to the congenital absence of one or more of the disaccharidases normally found in the mucosa of the small bowel When a patient with a congenital absence of a disaccharidase is fed the corresponding disaccharide he develops diarrhea with watery stools of low pH, and containing large amounts of the disaccharide as well as lactic acid Removing the specific disaccharide from the diet promptly causes the symptoms to disappear Due to the inability to absorb certain disaccharides by these patients, a *disaccharide tolerance test* can be done to demonstrate the disorder The test is similar to the oral glucose tolerance test and consists of giving a loading dose of the disaccharide and observing whether or not a rise in the blood glucose level occurs The interpretation of disaccharide tolerance tests must take into account the specificity of the various disaccharidases that have been identified in the small bowel mucosa of man Dahlqvist (1962) has listed the specificity of various disaccharidases, as shown in Table 2 11.

A congenital absence of lactase without intolerance to maltose or sucrose has been observed frequently This condition is usually recognized in newborns and infants who are unable to tolerate milk Sucrose and maltose intolerance are usually not recognized until later when these sugars are introduced into the diet Combinations of sucrose and isomaltose intolerance have

Table 2 11

Specificity of disaccharidases isolated from human intestinal epithelium and their quantitative importance for hydrolysis of different disaccharides

Enzyme		Substrate	Activity[a]
			%
Isomaltase	maltase Ia	Isomaltose	100
		Maltose	50
Invertase	maltase Ib	Sucrose	100
		Maltose	25
Maltase II		Maltose	15
Maltase III		Maltose	10
Trehalase		Trehalose	100
Lactase		Lactose	100
		Cellobiose	100

From Dahlqvist (1962).

[a] Total activity against each substrate

been recognized From a study of Table 2 11 it will be seen that four different maltases have been identified Thus, maltose intolerance is extremely rare, as it requires the absence of all four of these enzymes Such a patient would also be intolerant to isomaltose and sucrose

ABSORPTION OF PROTEIN

It has been estimated that approximately 225 g of protein may be absorbed per day by the small bowel of an adult man The protein available for absorption in the small bowel has both an *exogenous* and *endogenous* origin To maintain nitrogen balance, a 70-kg man requires approximately 50 g of dietary, or *exogenous*, protein per day The source of *endogenous* protein is protein contained in the gastrointestinal secretions and the desquamated epithelial cells of the gut The protein of the secretions may amount to 150 g /day, and that from desquamated cells 25 to 30 g /day Thus, the exogenous dietary protein is considerably diluted by the endogenous protein

Although some absorption of native protein and of intermediate products of protein digestion take place in the small intestine, the amount absorbed is insignificant from a nutritional standpoint The greater part of protein in the food is reduced to amino acids before being absorbed Protein absorption is, therefore, essentially amino acid absorption

All evidence indicates that amino acids are readily absorbed throughout the entire length of the small intestine They are taken up about as rapidly as they are produced, consequently, only small amounts of amino acids can be found at any time in the intestinal contents In fact, due to the dilution of exogenous protein previously mentioned, Nasset et al., (1955) have shown that feeding a dog a protein meal has little effect on the distribution of free amino acids in the intestinal contents that can be recovered from the lumen Any excess of amino acids in the portal blood due to absorption disappears rapidly Although all the amino acids are rapidly absorbed from the intestine, some are absorbed more rapidly than others The following list is in the order of rapidity of absorption glycine, alanine, cystine, glutamic acid, valine, methionine, leucine, tryptophane, isoleucine, norleucine, and isovaline (Verzar and McDougall, 1936)

Mechanism of Amino Acid Absorption. Wiseman (1953) demonstrated that the levorotatory, or naturally occurring forms of alanine, phenylalanine, isoleucine, valine, methionine, and histidine are absorbed more rapidly than the dextrorotatory isomers, and that the levoforms could be transferred through the intestinal wall against a concentration gradient, whereas the dextroforms tended to reach the same concentration on the two sides of the membrane (as did glutamic and aspartic acids) This and the fact that there are differences in the rate of absorption of amino acids that are not related to their diffusibility, as well as the rapid disappearance of amino acids from the intestine, all indicate that there is an efficient mechanism for the active absorption of many of the amino acids This is further supported by the fact that amino acid absorption is inhibited by the absence of oxygen, and by several metabolic inhibitors (Fridhandler and Quastel, 1955) Phloridzin, which interferes with the active absorption of glucose, does not interfere with active absorption of amino acids Nevertheless, amino acids are said to compete with sugars for the absorptive mechanism, that is, sugars are absorbed more slowly from a sugar-amino acid mixture than from a pure sugar solution (Cori, 1931)

At present, the mechanisms for the transport of amino acids may be classified as follows *1*) Neutral amino acids (including histadine) most efficient and operates at high concentrations *2*) a mechanism shared by the basic amino acids (ornithine, arginine, and lysine) Cystine may use this route in addition to pathway 1 *3*) Another pathway is shared by proline, hydroxyproline, sarcosine, N-dimethylglycine, and betaine The first two may use pathway 1 *4*) Proline and glycine may have an alternative pathway

A number of neutral amino acids compete with each other for the same transport system through the cell membrane The α-carbon of a neutral amino acid has four different groups attached to it a carboxyl group, an amino group, a hydrogen ion, and a side-chain radical, usually referred to as R An alteration in any of these groups attached to the α-carbon may interfere with the active transport of the amino acid Thus, when the charge on the carboxyl group is removed by the formation of an ester or an alcohol, a compound is produced that is not actively transported by the small bowel The removal of the charge of the amino group by acetylation, or by replacement of the amino group by a hydroxyl group, also results

in an inactive compound The α-hydrogen also has been shown to be essential for active transport to occur Replacing the α-hydrogen by a methyl group reduces the rate of transport A wide variety of side chains or radicals may be attached to the α-carbon and be actively transported Thus, the carrier system for neutral amino acids is not selective in regard to the side chain, as long as the side chain does not contain a charge If either a negative or positive charge is introduced into the side chain, active transport is interfered with

The levoforms of the basic amino acids lysine, arginine, and ornithine, are transported at rates of only about one-tenth of that of most of the neutral amino acids Like the neutral amino acids, the basic amino acids seem to share a common, but separate, transport system Thus, any two of the above-mentioned basic amino acids will inhibit the transport of the remaining basic amino acid in the group For some unknown reason, cystine, which is a neutral amino acid, appears to share the same transport system as the three basic amino acids

As previously mentioned, the transport of substances through the small bowel mucosa may involve a number of "steps" During absorption, amino acids accumulate in the mucosal cells and their intracellular concentration may exceed that in the lumen Thus, entry into the cells appears to be the active step With accumulation within the cell a favorable concentration gradient is created for diffusion out of the cell into the portal blood

Transamination During Absorption Most amino acids are not altered during transport through the epithelial cells lining the small bowel However, glutamic acid and aspartic acid may undergo transamination with pyruvic acid so that alanine is produced and released into the portal blood Thus, some amino acids are not necessarily delivered to the blood in the same form in which they are absorbed In the case of glutamic and aspartic acids, if the absorption rate is slow, practically all the glutamic and aspartic acids undergo transamination to alanine, on the other hand, during rapid absorption, there is less transamination, and considerable quantities of glutamic and aspartic acids appear in the blood (Neame and Wiseman, 1957) It has been suggested that in this way the intestinal mucosa can modify an amino acid mixture that is being absorbed from a particular meal, reducing the amount of amino acids that are in excess in the digesting mixture and increasing the amount of those which are in short supply and so more nearly meet the needs of the organism for specific amino acids

Absorption of Peptides, Proteoses, and Native Protein Many of the intermediate products of protein digestion are soluble in water and some are diffusible Thus, the question naturally arises as to whether peptides, proteoses and peptones may be absorbed from the intestine without being reduced to amino acids Peptone solutions certainly disappear from intestinal loops but this does not prove that the peptone is absorbed as such It will be recalled that the peptidases of the intestinal secretion are believed to be intracellular enzymes, which are not actually secreted, but appear in the intestinal contents as a result of a breakdown of shed epithelial cells It seems unlikely that peptides could get through the mucosal cell in any considerable quantity without being reduced to amino acids by the intracellular peptidases

This same reasoning does not apply to native protein, however, which may have escaped digestion by the pepsin of gastric juice or the proteolytic enzymes of pancreatic juice for one reason or another Except for traces of pepsin in the Brunner's gland secretion, there are no proteases in the intestinal secretion or in the mucosal cells which are known to attack native protein Native proteins are absorbed by the small bowel by a process similar to phagocytosis known as *pinocytosis* The absorption of native proteins by the process of pinocytosis has been demonstrated in the newborn of many mammalian species The ability of the small bowel mucosa to absorb protein in this fashion disappears, however, from a few days to a few weeks after birth, depending upon the species The mechanism is particularly valuable in those species that do not receive a passive immunity from the mother through the placenta In these species, passive immunity is induced by the absorption of immunoglobulins from the colostrum

The adult mucosa, in all mammalian species studied, is very resistant to the absorption of native protein except in antigenic amounts Thus, it is not surprising to find that egg albumin and milk protein can sometimes be demonstrated in the blood by immunological methods after these proteins have been placed in the intestine Absorption of native protein is

not necessarily a result of a pathological condition of the intestinal mucosa, although the absorption may be greatly increased by an inflammatory condition

ABSORPTION OF FAT

Dietary fat, consisting mainly of triglycerides, is insoluble in water, yet manages to be absorbed by the intestinal mucosa and appears in the thoracic lymph carried there as chylomicrons To accomplish this transfer, complex physicochemical reactions take place both in the lumen of the gut and in the intestinal columnar cell, in which the dietary triglyceride is hydrolyzed to monoglyceride, fatty acid, and glycerol and resynthesized back to triglyceride In this process, pancreatic lipase and bile salts play critical roles

Lumen Phase In the upper small bowel emulsified triglyceride is acted upon by pancreatic lipase This enzyme acts at the interface between the oil droplet of the triglyceride emulsion and the water phase in which it is suspended The emulsified triglyceride is hydrolyzed at the 1 and 3 position forming 2-monoglyceride and fatty acids

The monoglycerides and fatty acids then in the presence of bile salts form polymolecular aggregates termed "micelles," (already referred to in the section on bile salt formation and secretion) which also contain some cholesterol and phospholipid It is from these micellar aggregations that the lipids penetrate the columnar cell As a result, if the intestinal contents are sampled after a fatty meal, a clear and an oily phase can be distinguished Hofmann and Borgström (1966) have shown that the heavier clear phase contains fatty acids and monoglyceride, while the oily layer contains some di- and triglycerides

Penetration phase. The diameter of the micelle particles varies from 40 to 200 A It is thought that fat in this form is presented to the epithelial surface for absorption The exact mechanism whereby the fatty acids and monoglycerides penetrate the plasma membrane of the microvilli is not known However, the conjugated bile acids in the micellar dispersion are not absorbed in the upper small bowel at the site of the absorption of fatty acids and monoglycerides The conjugated bile acids are turned back to the lumen of the gut and are re-used in the formation of new micelles At present, the mechanism of penetration is presumed to be by diffusion, independent of temperature, and nonenergy requiring, and the available evidence certainly favors diffusion as the primary mechanism

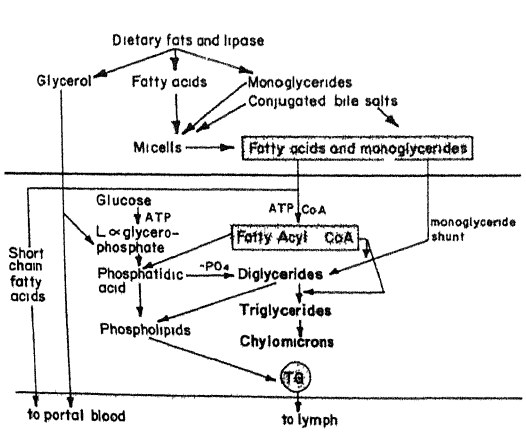
Intracellular phase This phase concerns the re-esterification of fatty acids and monoglycerides to restore triglycerides This is accomplished by the migration of monoglycerides and fatty acids to the endoplasmic reticulum, where fatty acids are activated to the coenzyme A (CoA) derivatives and resynthesized to triglyceride by the acylation of monoglycerides in a multi-enzyme complex

Resynthesis of fatty acids into triglycerides within the epithelial cells has been abundantly confirmed Bergstrom and others (1954) found that when free oleic acid labeled with ^{14}C was fed, 63% of the fed acid was recovered in the lymph as neutral fat When triolein was fed, 65% was recovered in the lymph These authors also found that when fatty acids containing 10 or fewer carbon atoms were fed, they were absorbed by way of the portal blood, where they were transported mainly as free fatty acids Longer-chain fatty acids were synthesized in the epithelial cells into triglycerides and were absorbed by way of the lymphatics Kiyasu, Bloom, and Chaikoff (1952) also found that short-chain fatty acids were transported mainly by the portal pathway and long-chain saturated fatty acids by way of the lymph Therefore, a sort of partition does occur in the intestine

It is now recognized that one of the major functions of the small bowel is the resynthesis of triglycerides from fatty acids and monoglycerides, and packaging them in the form of *chylomicrons* The current state of knowledge regarding the intracellular synthesis of triglycerides has been recently reviewed by Johnson (1968) The synthesis of triglyceride requires energy which is supplied in the form of adenosine triphosphate (ATP) from glycolysis or oxidative phosphorylation In addition, magnesium ions and coenzyme A (CoA) are necessary The fatty acids are first "activated" by the formation of a CoA derivative, which not only makes the fatty acids more reactive, but also water soluble As a result of glucose breakdown, glycerophosphate is formed The glycerophosphate combines with the activated fatty acids to form phosphatidic acid, which after loss of its phosphate group becomes a diglyceride The diglycerides are then esterified further to triglycerides, or converted to phospholipids

These events are shown diagrammatically in Figure 2-45. The direct esterification of absorbed monoglycerides to diglycerides has been suggested, and has been referred to as a "monoglyceride shunt." Senior and Isselbacher (1962) have obtained evidence for the direct esterification of monoglycerides and have demonstrated the presence of a monoglyceride lipase within the epithelial cell. This lipase is distinct from pancreatic lipase and can split β-monoglycerides equally or even more readily than α-monoglycerides. It is relatively inert towards di- and triglycerides. Thus, if activated fatty acids are not available for combining with monoglycerides, the monoglycerides can be hydrolyzed into glycerol and a fatty acid. If the fatty acids are not available for combining with available monoglycerides, the biosynthesis of triglyceride can occur via the α-glycerophosphate pathway. In the resynthesis process, it is now held that the free glycerol liberated in the intestine is available for incorporation.

The final step in the synthesis of triglycerides by the epithelial cells is the formation of chylomicrons for transport in the lymph. The chylomicrons are covered by a fine protein surface coating, amounting to less than 1% of the weight of the chylomicron. While the triglycerides are being synthesized, synthesis of phospholipid and protein is also occurring within the columnar cell, which, in combination with cholesterol esters, forms a β-lipoprotein which is the coating referred to. There is electron microscopic evidence that the chylomicrons are extruded from the lateral margins of the epithelial cells by a process analogous to the pancreatic secretion of zymogen granules. The chylomicrons leave the epithelial cells at about the level of the nucleus, passing into the intercellular spaces. They eventually pass across the basement membrane and into the lymphatic vessels by passage between the endothelial cells.



Short-chained fatty acids absorbed by the epithelial cells are passed to the portal blood for transport to the liver. The fate of glycerol hydrolyzed in the lumen of gut during the process of digestion has been debated. From early isotope studies, Reiser and his co-workers (1952) concluded that glycerol released in the lumen of the gut during the process of digestion was not utilized in the resynthesis of triglycerides within the epithelial cells, but apparently took some other metabolic pathway. Recently, Saunders and Dawson (1962) have demonstrated that, if labeled glycerol is administered with a bile salt-fatty acid mixture, considerable incorporation of the label into lymph triglyceride glycerol occurs. Thus, apparently two sources of glycerol are available to the epithelial cell for the synthesis of triglycerides, that absorbed directly from the intestinal lumen and that available from the conversion of glucose to glycerophosphate. Hogben (1960) has pointed out that the observed absorption of fatty acids against a concentration gradient may be due to differences in pH on each side of the membrane as a consequence of ion transport rather than some "active" process which is specific for fatty acids.

Site of Fat Absorption. Fat absorption takes place most rapidly in the upper intestine, particularly in the duodenum and upper jejunum. In man, Borgström and his associates (1957) have demonstrated that fat absorbed mainly from the upper small bowel-duodenum and proximal jejunum. They fed a measured quantity of fat in a meal containing a nonabsorbable marker, polyethylene glycol, and sampled the intestinal contents at different levels. By comparing the concentration of intraluminal fat to the concentration of polyethylene glycol, they were able to calculate fat absorption at different levels of the small bowel, and found that practically all of the fat had been absorbed by the time the meal had reached the distal jejunum.

Fat absorption is impaired following bilateral adrenalectomy, due to the absence of cortical hormones. The role of the cortical hormones in

fat absorption has not been clarified but it has been suggested that they modulate the enzymes involved in the synthesis of triglycerides in the mucosal cells

ABSORPTION OF CHOLESTEROL

Normally, about a half a gram of cholesterol is consumed each day in the diet This dietary cholesterol is mixed with 2 to 3 g of endogenous cholesterol within the lumen of the small bowel The source of the endogenous cholesterol is the *bile* which may contain cholesterol in the amount of 1 5 to 2 5 g /day, and from *desquamated epithelial cells* which may contain up to half a gram of cholesterol per day Thus, the small bowel normally absorbs $2^1/_2$ to 3 g of cholesterol per day Only about one-fifth or one-sixth of this amount is represented by the dietary cholesterol Borgström (1960) has demonstrated that cholesterol absorption in man occurs at higher rates in the upper small bowel than in the lower small bowel He used a cholesterol labeled with radioactive carbon and determined its rate of disappearance from various levels of the small bowel It has long been known that cholesterol absorption differs from fat absorption in a number of important respects Only 40 to 60% of dietary cholesterol is absorbed, as opposed to the absorption of 95% of dietary fat When radioactive cholesterol is administered as a single dose orally, and the radioactivity of the venous blood is determined, it is found that the peak radioactivity does not appear until 2 to 3 days later (Biggs et al , 1952) When a similar study is done using radioactive fat, peak blood radioactivity is usually found 3 to 4 hr after the oral dose Although cholesterol probably enters the intestinal epithelium at a rate comparable to that of fat, once inside the cell, it is accumulated and released very slowly This accounts for the delay in appearance of labeled cholesterol in the blood

Bile is essential for cholesterol absorption, and pancreatic juice also plays an important role Fats in the diet appear to stimulate cholesterol absorption Vahouny (1958) and coworkers found that optimal absorption of cholesterol occurred from a mixture of fatty acid, taurocholate, and cholesterol Fatty acids markedly stimulate the absorption of cholesterol The stimulatory effect of triglycerides may be due to the fact that they provide a source of fatty acids through the hydrolysis of fat by pancreatic lipase

Most of the cholesterol in the diet is in the form of esters, but in the small bowel these esters are hydrolyzed by a cholesterol esterase present in pancreatic juice Cholesterol esters are not absorbed as readily as free cholesterol However, once absorbed, cholesterol is again esterified, and approximately two-thirds of the cholesterol appearing in the lymph is in the esterified form As dietary cholesterol esters are being hydrolyzed to free cholesterol, dietary triglycerides are being hydrolyzed to monoglycerides and fatty acids Under proper conditions, these substances (bile salts, phospholipids, monoglycerides, and fatty acids) form mixed micelles containing free cholesterol The free cholesterol within the mucosal cell is esterified in major portion and transferred to the lymph along with triglyceride, phospholipids, and protein in the form of chylomicrons Figure 2 46 diagrams the current concept of cholesterol absorption

ABSORPTION OF VITAMINS

The *water soluble vitamins* are assumed to be absorbed by the intestinal mucosa, along with

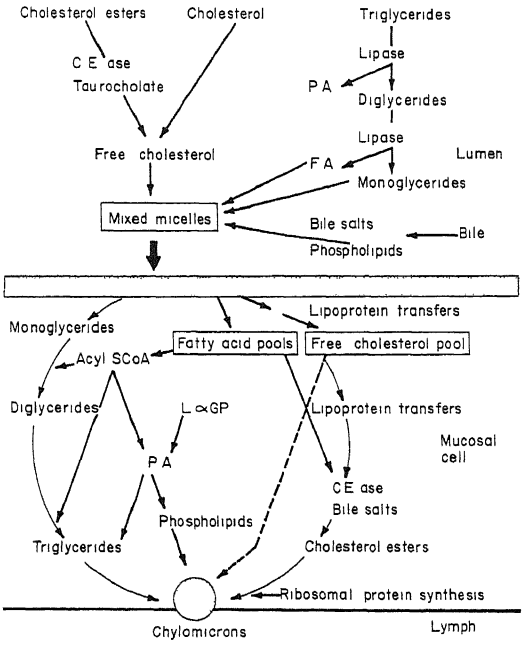


Fig 2 46 Mechanism of cholesterol absorption Abbreviations C E ase cholesterol esterase F A fatty acids Diglyc diglycerides Monoglyc monoglycerides L ∝-GP L-∝-glycerolphosphate P A phosphatidic acid (From Treadwell and Vahouny 1968)

the other constituents of the foods in which they occur by passive diffusion and by the mechanisms which facilitate water movement, but this may not be true for folic acid It is generally believed that thiamin is less readily absorbed by older people than by young people in good health This may be applicable also to some of the other vitamins

Absorption of *fat soluble vitamins*, A D and K, probably takes place under the same conditions as the absorption of fat In any disease or dietary deficiency in which fat absorption is inadequate, the absorption of these vitamins is apt to be impaired This is demonstrated by the deficient absorption of vitamin K in obstructive jaundice A deficiency of this vitamin causes inadequate prothrombin production by the liver and a tendency to hemorrhage This situation is remedied by the parenteral administration of vitamin K Absorption of vitamin E is also said to be impaired in the absence of bile

Vitamin A is absorbed chiefly in the form of esters of the vitamin A alcohol with fatty acids The esters are hydrolyzed in the intestinal lumen but are re-esterified in the epithelial cells This mechanism parallels the one that is operative in the absorption of fat, it is not surprising, therefore, to find that the conditions that favor fat absorption also favor absorption of vitamin A The presence of bile in the intestine apparently is necessary for the absorption of carotene, a precursor of vitamin A, but not for vitamin A itself The absorption of the fat-soluble vitamins is adversely effected by the presence in the intestine of unabsorbable oil, such as mineral oil

Intestinal Absorption of Vitamin B_{12}. Both passive and active mechanisms exist for the absorption of vitamin B_{12} The passive mechanism probably diffusion, occurs throughout the entire small intestine, and is important only in the presence of pharmacological amounts of vitamin B_{12}, whereby approximately 1% of free vitamin B_{12} is absorbed The active mechanism is intrinsic-factor mediated, operative only in the ileum and of primary importance in the absorption of physiological (2 μg or less) amounts of vitamin B_{12} Intrinsic factor (IF) a glycoprotein secreted by the parietal cells of the stomach forms a complex with the free vitamin B_{12} molecule released from food and facilitates the delivery of vitamin B_{12} to specific ileal receptor sites and also protects vitamin B_{12} from intestinal enzymatic degradation and from bacterial utilization The attachment of the IF-B_{12} complex to the ileum requires an intact ileal receptor is pH dependent, and requires the presence of divalent cations (Ca^{++} and/or Mg^{++}) The nature of the IF-B_{12} ileal receptor is unknown but it appears to be particulate and either a glycoprotein or mucopolysaccharide The fate of IF once the IF-B_{12} complex attaches to the ileal mucosal cell is unknown, but most evidence suggests that IF is not absorbed and that the B_{12} is released prior to entering the mucosal cell Little is known of the intracellular pathway of the B_{12} molecule but the actual uptake is temperature, oxygen, and glucose dependent Once vitamin B_{12} traverses the lining of the intestine, it enters the portal blood bound to a carrier serum protein which is distinct from IF

9/MOVEMENTS OF THE ALIMENTARY CANAL

Disturbances of motility of the alimentary canal account for more symptoms of gastroenterological disorders than alterations in absorptive or secretory mechanisms In fact, most of the major diseases of the digestive tract are manifested by a disturbance of motor activity Dysphagia, vomiting, retching, crampy abdominal pain, diarrhea, and constipation are examples of such symptoms

Mastication

The first mechanical process to which the food is subjected in its progress through the alimentary tract is mastication, or chewing, which serves to break down the food particles into smaller masses, to enable the saliva to lubricate and moisten dry food, and to distribute the salivary constituents throughout the food mass The decrease in the size of the food particles is accomplished by the action of the teeth through the movements of the lower jaw The moistening is accomplished by the addition of saliva to the food, and lubrication is due to the mucus secreted by the salivary glands Starch digestion in the mouth by the salivary amylase is unimportant, but the presence of food in the mouth gives rise to stimuli which serve to initiate reflexes that are important in the later stages of digestion

The movements of the lower jaw in chewing are brought about by the muscles of mastication, which include the masseter, the temporal muscles, and the internal and external pterygoid muscles, all of which receive their nerve supply from the mandibular division of the trigeminal nerve The buccinator muscle, which may also be considered a muscle of mastication, is supplied by the facial nerve Through the action of the jaw muscle, the mandible may be moved in any one of several directions Simple closing of the mouth, as in biting or crushing food, is accomplished by the combined action of the masseter and temporal and internal pterygoid muscles Opening the mouth is accomplished by the action of the relatively weak digastric and myloid muscles and is aided by gravity The grinding movements of the molar teeth are accomplished by a rotational movement of the jaw This kind of chewing is accomplished by the alternate contraction of the right and left pterygoid muscles When all four pterygoid muscles contract, protrusion of the jaw results Retraction of the jaw is accomplished by contraction of the posterior fibers of the temporal muscles and the geniohyoid muscles

The force of the bite is limited by the sensitivity of the peridental membrane and not necessarily by the power of the muscles of mastication In man, the molars have been observed to exert a pressure in excess of 270 lb The pressure exerted by the incisor teeth is less, due to the lesser mechanical advantage in this location According to Howell and Manly (1948), pressure exerted by the incisors ranges from 11 to 25 kg These authors found a force of 29 to 90 kg for the molars A dog, when chewing bones, may exert a force as great as 165 kg (Triska, 1924)

Chewing may be carried on voluntarily, but for the most part it is a reflex act The dual nature of the nervous mechanism is indicated by the fact that chewing movements may be elicited by electrical stimulation of appropriate areas of the cerebral cortex

Deglutition

Magendie (1847) described the act of swallowing as occurring in three stages which he described as follows "For the better understanding of the subject we shall divide deglutition into three stages In the first the aliments pass from the mouth into the pharynx, in the second they pass over the opening of the glottis and nasal fossae and arrive at the esophagus, in the third they pass through this canal into the stomach " It is convenient to refer to Magendie's first stage as the oral stage, the second stage as the pharyngeal stage, and the third as the esophageal stage

NEURAL CONTROL

The Swallowing Reflex The first stage of swallowing, like chewing, may be initiated voluntarily but is normally a reflex action The second and third stages of swallowing are entirely involuntary If one wishes to swallow when the mouth is free of food or foreign material, a little saliva is passed backwards by the tongue and thus serves as a mechanical stimulus for the initiation of the second and third phases of the act which are purely reflex If the mouth is kept perfectly free of saliva, deglutition becomes impossible

Like chewing, swallowing may be produced by stimulation of appropriate areas of the cerebral cortex or by stimulation of local receptors, even in the decerebrate preparations The receptors for initiation of the swallowing reflex are situated in the vicinity of the entrance to the oral pharynx in all animals but are distributed somewhat differently in different species

In man, the receptors for the swallowing reflex are distributed in a ringlike fashion around the entrance to the oral pharynx and are found in the mucous membrane covering the anterior and posterior pillars of the fauces, the tonsils, the soft palate, the base of the tongue, and the posterior pharyngeal wall According to Pommerenke (1928), the most sensitive areas in the human are in the vicinity of the anterior and posterior pillars of the fauces and the tonsils

The Swallowing Center Coordination of the numerous somatic and visceral muscles involved in the act of swallowing is controlled by a group of neurons in the floor of the fourth ventricle known as the swallowing center, first described by Markwald in 1889 He found that destruction of the brain substance above and somewhat lateral to the alae cinereae of the medulla eliminates the swallowing reflex while leaving respiration intact Destruction of the alae cinereae stops respiration but does not interfere with swallowing Thus, the swallowing center is closely associated with, although distinct from, the respiratory center This fact is of considerable clinical interest, because in diseases of the medulla, such as bulbar polio or brain tumor, difficulty in swallowing often precedes respiratory failure

Innervation of the Muscles of Deglutition. The efferent nerve fibers involved in the act of swallowing are chiefly in the hypoglossal, glossopharyngeal and vagus nerves. The hypoglossal and glossopharyngeal nerves are chiefly concerned with the buccopharyngeal stage, the vagus with the esophageal stage of swallowing The vagus nerves supply the striped muscle of the esophagus with motor fibers, hence, one of the more serious consequences of cutting the vagi above the origin of the main esophageal branches is interference with swallowing The striped muscle of the esophagus is permanently paralyzed after a high vagotomy (e g , in the neck) but the smooth muscle of the lower third shows some degree of recovery, peristalsis in this portion may return and swallowing may again become possible In animals such as the dog and rabbit in which the entire esophagus is made up of striped muscle, high vagotomy results in permanent inability to swallow Paralysis of the esophagus, with the resultant accumulation of food and liquid in the esophagus and pharynx and associated with the loss of the sensory innervation of the larynx results in the aspiration of food and liquids into the lungs, the animal either suffocates or dies of aspiration pneumonia Aspiration pneumonia is the most common cause of death after a high vagotomy in dogs

The striped muscle is directly innervated by the vagus nerves, the fibers ending on the muscle cells without the intervention of any local ganglion cells In the smooth muscle portion, however, there is an extensive local plexus including nerve cells and their branches similar to the myenteric plexus in other parts of the gastrointestinal tract In this part the vagus fibers do not go to muscle cells but end on nerve cells within the esophageal plexus

THE FIRST OR ORAL STAGE OF SWALLOWING

The first stage of deglutition consists of passage of material through the oral cavity into the pharynx This stage of swallowing is initiated voluntarily With solid food, the act is usually preceded by mastication The movements of the tongue against the hard palate and the cheeks help to form a bolus by moistening and mixing the material with saliva, thus softening and lubricating the bolus for passage through the pharynx and the esophagus After this intraoral preparation, the bolus is maneuvered to a position on the posterodorsal surface of the tongue Bosma (1957) has described this as the "preparatory position," as he has observed that the bolus is always maneuvered to a constant position on the surface of the tongue

With the bolus in the preparatory position, swallowing is initiated The front portion of the tongue is retracted and depressed, the hyoid is elevated slightly, mastication ceases, respiration is inhibited reflexly, and the back portion of the tongue is elevated and retracted against the hard palate This latter action of the tongue is produced principally by the myelohoid, styloglossus and hypoglossus muscles, which force the bolus into the pharynx Intraoral pressures recorded during swallowing have demonstrated that in the anterior part of the mouth near the incisor teeth, negative pressure changes as great as minus 300 cm of water may occur (Miyakawa et al , 1957) In the posterior part of the oral cavity, positive pressure changes occur due to the forceful contraction of the tongue against the hard and soft palates Positive pressures as great as 100 cm of water are not uncommonly recorded in the posterior part of the oral cavity during swallowing (Hightower, 1958)

Disturbances of the first stage of swallowing Disturbances may be caused by a number of conditions Any painful inflammatory process involving the mouth or tongue may make swallowing difficult Neoplasms of the tongue or pharyngeal strictures may obstruct the passage of the bolus into the pharynx Paralysis of the tongue can interfere markedly with swallowing Congenital anomalies of the lips, tongue, and palate are commonly associated with difficulty in initiating the act of swallowing

THE SECOND OR PHARYNGEAL STAGE OF SWALLOWING

The contact of material from the mouth with the pharyngeal and peripharyngeal structures initiate reflexes that complete the second and third stages of deglutition Anesthetization of these structures inhibit the swallowing reflexes The swallowing act, once initiated, is dominant over other functions occurring in this area Thus, respiration and speech are interrupted by the second stage of the swallowing act The complexity of the act, and the high degree of integration necessary, is appreciated when one recalls the timed sequence of events in this stage of deglutition and that the second stage is completed in less than 1 sec

Understanding of the second stage of swallowing will be easier if we keep in mind the fact that the alimentary tract crosses the respiratory tract in the region of the pharynx, so the swallowing movements have to serve two quite different functions one, to propel the food through the pharynx and into the esophagus, and the other, to protect the airway, both above and below, from the possible entrance of food particles Although the propulsive and protective mechanisms operate simultaneously it is simpler to describe them separately

When the bolus reaches the base of the tongue, its movement is accelerated by a rapid downward and backward movement of the tongue which takes place at the same time as the forward movement of the larynx With the entrace of the bolus into the larynx, contraction of the superior pharyngeal constrictor occurs, initiating a rapid pharyngeal peristaltic wave which moves down the pharynx, propelling the bolus in front of it The walls and structures of the hypopharynx are elevated to engulf the oncoming bolus The cricopharyngeus muscle, which has kept the esophagus closed until now, relaxes as the bolus approaches and allows it to enter the upper esophagus The peristaltic contraction continues into the esophagus to become the primary esophageal peristaltic contraction

As a result of the combined forces acting on it, the bolus moves with such speed that it is often projected deep into the esophagus Rushmer and Hendron (1951) using cinefluoroscopy to study pharyngeal transport, observed that a bolus approached the pharynx with the velocity of 9 ft /sec , passing through the pharynx, it attained a velocity of 19½ ft , 24 ft , and 28½ ft /sec in successive frames, then slowed to 12 ft /sec , and finally to 5 ft /sec as it passed into the esophagus. These data were obtained with the subject in a horizontal position, in the erect position, the bolus moved even faster

We must now consider how the airway is protected while the food is passing the pharyngeal crossroads It will be recalled that there are four possible outlets from the oral pharynx through which food may be expelled These are back into the mouth, up into the nasopharynx, forward into the larynx, and downward into the esophagus The swallowing reflex has to be so coordinated that the food takes only one of these possible paths, namely, into the esophagus Return into the mouth is prevented by the position of the tongue against the roof of the mouth and the high pressure developed in this area Passage into the nasopharynx is normally prevented by the combined action of the tensor veli palatini and the levator veli palatini

muscles which together stiffen the soft palate and cause it to press against the posterior pharyngeal wall. It thus forms a protective partition between the oral pharynx and the nasopharynx. If these muscles are paralyzed or if there should be a defect in the soft palate, food enters the nasopharynx and swallowing becomes extremely difficult or impossible. An important factor in preventing entrance of food into the respiratory tract during swallowing is the inhibition of respiration which occurs during the second stage of swallowing (Magendie, 1847).

The laryngeal opening is closed during the act of swallowing by the approximation of both the true and false vocal cords. In addition, the epiglottis serves to divert the oncoming bolus to one side or the other of the larynx. Only when liquid is being swallowed in considerable volume does the swallowed material cascade over the end of the epiglottis. Most observers agree that the epiglottis is carried into a horizontal position by the backward movement of the tongue and the forward movement of the larynx. This same combination of movements results in the larynx being drawn up under the base of the tongue where it is completely out of the way of the oncoming bolus.

Disturbances of the second stage of swallowing. Disturbances of the second stage of swallowing are not unusual. Acute pharyngitis is probably the most common cause of painful dysphagia of the second stage of swallowing. Tonsillitis is another example. Pharyngeal paralysis is indicative of serious neurological disease and dysphagia developing in poliomyelitis, syringomyelia, multiple sclerosis, or after a cerebrovascular hemorrhage points to bulbar involvement. In these patients, it may be impossible for the soft palate to be elevated, thus allowing liquids to be regurgitated through the nose. The pharyngeal constrictions may be unable to initiate the peristaltic contraction which empties the contents of the pharynx into the esophagus. Cranial nerve (glossopharyngeal) neuritis also may result in pharyngeal paralysis. Diptheria is sometimes followed by palatine and pharyngeal paralysis.

Mobility of the larynx is essential to the second stage of deglutition. Any disease process that prevents the larynx from ascending to lie beneath the base of the tongue causes marked dysphagia. Neoplasms, such as carcinoma of the larynx and thyroid, and infectious conditions, such as tuberculosis and syphilis, may result in fixation of the larynx, thus making swallowing extremely difficult.

THE THIRD OR ESOPHAGEAL STAGE OF SWALLOWING

At the junction of the esophagus with the pharynx above and with the stomach below, there exist physiological sphincteric mechanisms that effectively maintain and separate the slightly negative intraluminal pressure of the esophagus from the positive pressures of the pharynx and stomach. Except during deglutition these esophageal sphincters are normally closed and prevent the aspiration of air from above, or gastric juice from below, into the esophagus. They respond to deglutition by first relaxing and then contracting in a sequential manner with the primary peristaltic wave as it passes over these junctional zones. In describing the third stage of deglutition, the motor activity of the esophageal sphincters will be considered separate from that of the body of the esophagus.

The Upper Esophageal Sphincter. The pharyngeal junction is characterized by a short zone of high-resting intraluminal pressure in the region of the inferior pharyngeal constricture or cricopharyngeus muscle at the level of the cricoid cartilage (Fyke and Code, 1955). The circular muscle fibers in this area remain in a stage of tonic contraction, producing this zone of high-resting intraluminal pressure, and effectively separating the pharyngeal and esophageal luminae. The zone is approximately 4 cm in length and usually is located between 15 and 20 cm from the incisor teeth. The basal-resting pressure in the high pressure zone varies from 20 to 30 cm of water above atmospheric pressure (Fig. 2.47).

When intraluminal pressures are recorded from the upper esophageal sphincter during deglutition, the pattern of pressure change is distinctly different from that recorded from the pharynx or upper esophagus (Fig. 2.48). Immediately after the onset of deglutition, usually within 0.5 sec, a brief increase in the high resting pressure first occurs, but this lasts only a few tenths of a second, to be followed by a marked decrease in pressure below the resting basal pressure, and usually negative in comparison with atmospheric pressure. The decrease in pressure in the pharyngoesophageal junction occurs almost simultaneously with the increase

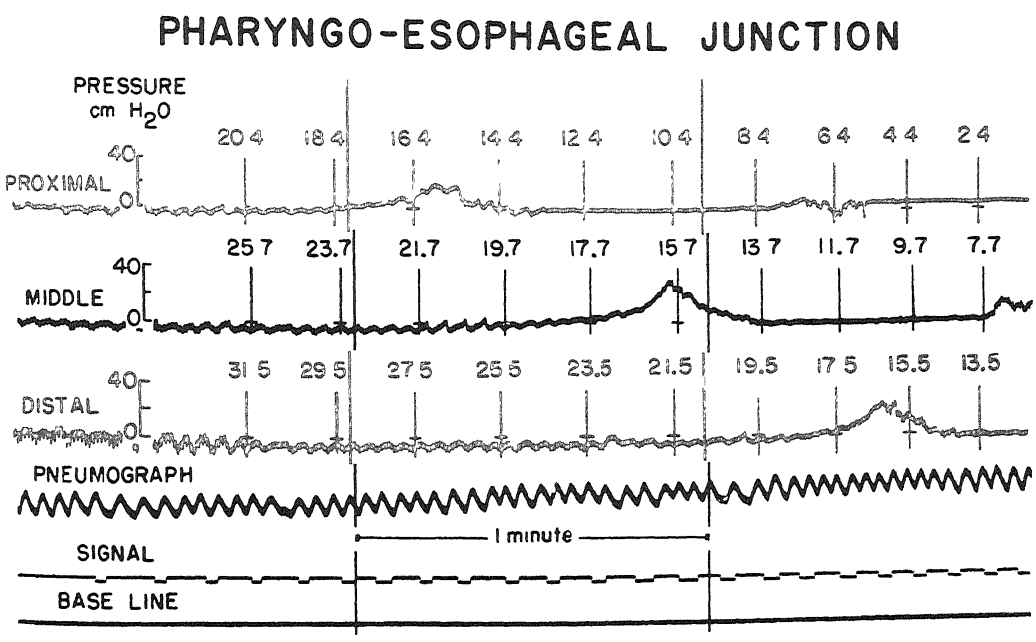


Fig 2 47 Demonstration of the zone of high pressure at the pharyngoesophageal junction The catheters were placed in the esophagus and slowly withdrawn through the upper esophageal sphincter while recording pressure continuously Each centimeter of withdrawal is indicated by the signal and the vertical lines indicate the distance in centimeters from the incisor teeth to the tip of each catheter The pull through technique clearly demonstrates the negative basal pressure in the esophagus and the zone of high pressure extending from 14 to 17 cm in the upper esophageal sphincter (From Hightower 1958)

in pressure in the pharynx produced by the peristaltic wave resulting from the contraction of the superior and middle pharyngeal constrictors These events markedly increase the pressure gradient from the pharynx across the upper esophageal sphincter With the arrival of the swallowed bolus at the pharyngoesophageal junction, the previously relaxed fibers of the upper sphincter contract and propagate the peristaltic contraction and pressure gradient into the upper esophagus

The small initial positive pressure wave observed in the pharyngoesophageal junction with deglutition may be due to an increase in the tonic contraction of circular fibers of the cricopharyngeus muscle, or it may be related to the elevation of the larynx and increased tension in this area The duration of the negative component of the junctional pressure pattern of deglutition is usually 0 7 to 1 0 sec , after which time the pressure has returned to the previous basal level The pressure continues upward, however, and a positive pressure wave, forming the third component of the pattern is observed This positive pressure wave represents a continuation of the pharyngeal peristalsis through the pharyngoesophageal junction The maximum

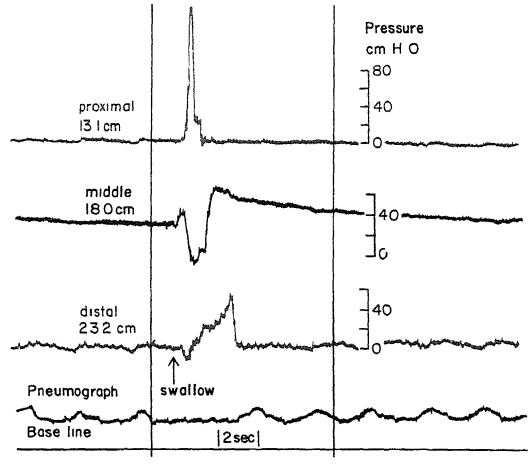


Fig 2 48 Pressure changes in the upper esophageal sphincter during deglutition The proximal catheter is recording pressure from the lower pharynx the middle one from the sphincteric area and the distal one from the upper esophagus The sequential appearance of peak pressures in each complex from above downward indicates the passage of the peristaltic contraction from the pharynx over the pharyngoesophageal junction into the upper esophagus (From Hightower 1958)

pressure produced by this wave is usually 70 to 90 cm of water above atmospheric pressure The pressure within the junctional zone then

returns to its previous basal level within 3 to 4 sec

Motility of the Body of the Esophagus

Primary peristaltic contractions previously described as originating in the pharynx and passing over the pharyngoesophageal junction continue into the esophagus sweeping the bolus onward and maintaining the pressure gradient The pressure pattern consists of an initial negative wave followed by three positive pressure components (Fig 2 49) The initial negative component is of brief duration and occurs immediately after the onset of deglutition It is observed more frequently in the upper than in the lower esophagus, and is thought to be due to the sudden stretching of the closed esophagus by elevation of the larynx The negative component is followed by an abrupt increase in pressure of 10 to 15 cm of water The third component consists of a slowly increasing positive pressure change or else a plateau of positive pressure is maintained The fourth and most prominent component of the deglutition pressure pattern is a large, simple, monophasic positive pressure wave that rises rapidly to a peak pressure and declines with equal speed The abrupt rise in intraluminal pressure following the initial negative component has been attributed to the sudden injection of the bolus or liquid into the esophagus, as it is observed to occur simultaneously at different levels of the esophagus The gradual increasing or plateau of pressure is related to the approaching peristaltic contraction Its duration and time is directly proportional to the depth of the recording catheter in the esophagus The large monophasic positive pressure wave that terminates the complex is recorded as the peris

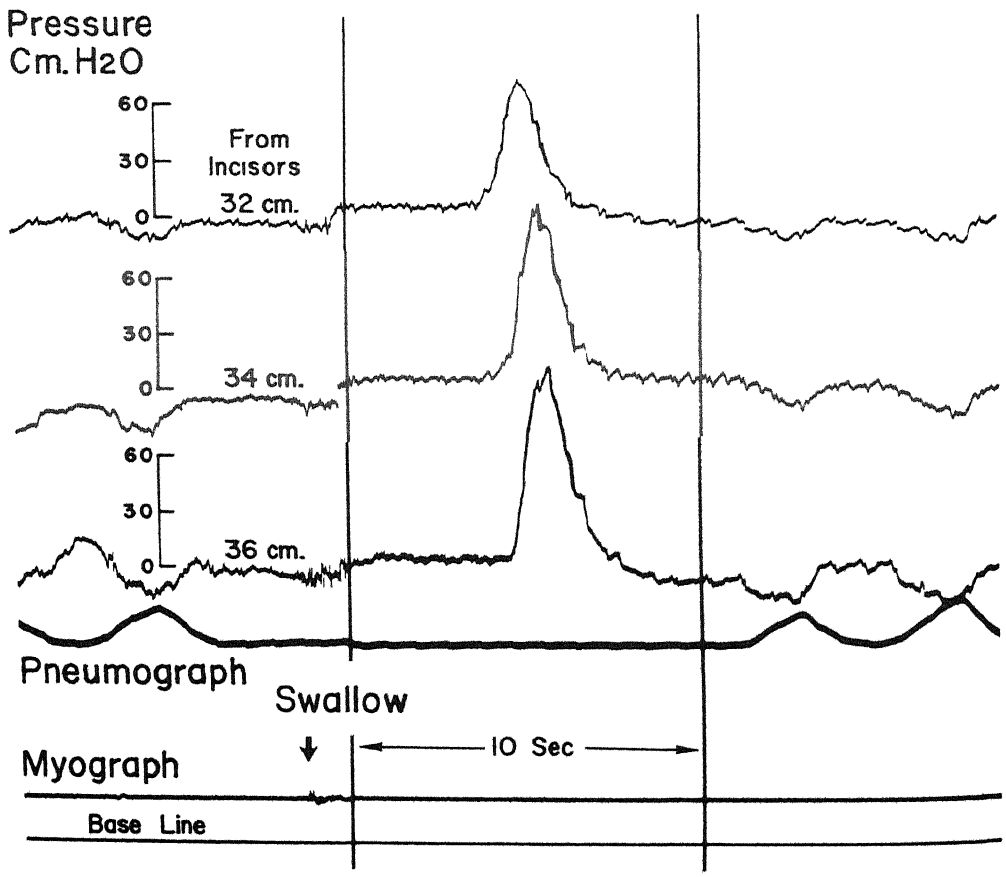


Fig. 2 49. Normal deglutition pressure complexes recorded from the lower esophagus See text for detailed description. The pneumograph is recording respiration Note that respiration was voluntarily stopped at end expiration a few seconds prior to deglutition The myograph records the action potentials of the accessory muscles of deglutition in the neck and indicates the precise onset of swallowing

taltic contraction passes over the open tip of the catheter

In man, the *rate of propagation* of the primary peristaltic contraction has been determined with the subjects in the horizontal position (Hightower and Salem, 1958) The time required for the primary esophageal peristaltic contraction to traverse the entire esophagus is from 8 to 12 sec As the peristaltic wave descends the esophagus, its rate of propagation becomes progressively slower It has been found that the rate of propagation of a primary contraction is quite rapid in the pharynx and upper esophagus Soon after it is initiated, it is traveling at a rate of approximately 50 cm /sec The rate rapidly decreases, however, as shown in Figure 2 50 When the peristaltic wave has traveled to a level 25 to 30 cm from the incisor teeth, its speed has decreased to about only 3 cm /sec For some unknown reason, the rate then increases slightly as the contraction ring approaches the esophagogastric junction After reaching the esophagogastric junction, the rate markedly decreases again until it is moving at less than 1 cm /sec

In the upright position, liquids travel through the esophagus at a speed that greatly exceeds the rate of propagation of the primary peristaltic wave This rapid transport is considered to be due to the propelling force created by the contraction of the pharyngeal constrictors during the second stage of deglutition, as well as by the action of gravity upon the swallowed liquid In the pharynx, near the end of the second stage of deglutition, pharyngeal pressure is mounting due to contraction of the pharyngeal constrictors and when the cricopharyngeus muscle suddenly relaxes and opens the upper end of the esophagus, which has a negative intraluminal pressure, the liquid is then more or less squirted into the lumen of the esophagus In the upright position, the rapid passage of liquid through the esophagus requires only a second or two Fluoroscopy demonstrates that swallowed liquids usually are arrested at the lower end of the esophagus and await the arrival of the peristaltic contraction before they are admitted into the stomach At times, however, when a series of swallows occurs rapidly, as when drinking a glass of water, the liquid may pass on through the esophagogastric junction with little pause and enter the stomach within a second or two after drinking has begun

Secondary peristaltic contractions are also observed in the esophagus These contractions arise locally as a result of distention, and progress down the esophagus in a manner similar to the primary contraction Secondary peristaltic contractions produce a pressure wave similar to the final pressure wave of the deluti tion pressure pattern It is a simple monophasic positive pressure wave that rises rapidly to a peak then returns promptly to the base line Secondary waves are usually of less amplitude than primary waves In the normal esophagus, secondary peristaltic contractions may be caused by local distention by fluid or a solid bolus If the primary wave is unable to propel the swallowed material, secondary contractions originate and complete the transport of the substance

Tertiary esophageal contractions also have been observed in the esophagus These contractions occur irregularly and locally, and are observed most frequently in the lower esophagus in elderly individuals They are not peristaltic

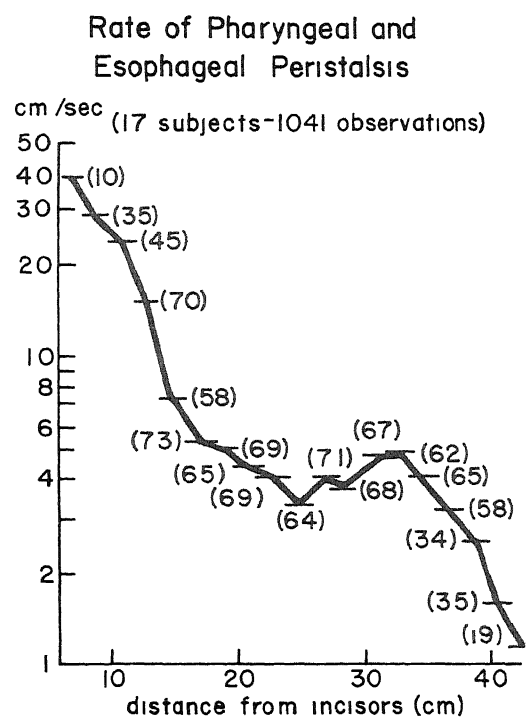


Fig 2 50 The rate of propagation of primary peristaltic contractions in the pharynx and esophagus as determined in 17 healthy adults The number of observations at each level is indicated in parentheses The short horizontal line indicates the distance over which the groups of observations were made

Their function is not known. Roentgenologically, they may produce a varied picture with a serrated or beaded appearance or like multiple diverticula-like pockets. The roentgenological picture of tertiary contractions has been referred to as "curling" and the "corkscrew esophagus." Lerche (1950) has suggested that contraction of the spiral muscular coat of the lower esophagus accounts for the tertiary contractions.

The Lower Esophageal Sphincter. Probably no other area of the alimentary canal has caused more confusion and diverse opinion than the terminal few centimeters of the esophagus. Only within the past few years have anatomists, physiologists, radiologists and endoscopists been able to reach some common agreement and understanding of the structure and function of the esophagogastric junction. It was Lerche's (1950) classic studies correlating anatomic and radiological observations that stimulated reinvestigation of this area. Sanchez, Kramer, and Ingelfinger (1953) demonstrated that the distal 2 to 5 cm. of the esophagus has a characteristic motor function. They found that the characteristic deglutition pressure pattern observed in the upper esophagus was not propagated into the vestibule. In addition, these same authors later demonstrated that the body of the esophagus and the vestibule react inversely to cholinergic and anticholinergic stimulation (Ingelfinger et al., 1954).

Fyke, Code, and Schlegel (1956) explored the esophagogastric junction of normal persons with a miniature pressure transducer and found a zone of high pressure similar to that encountered at the pharyngoesophageal junction. The high pressure zone extended from about 1 to 2 cm. below the hiatus of the diaphragm to 1 to 2 cm. above the hiatus. The mean resting maximum pressures in this region, expressed as centimeters of water above the pressure in the fundus of the stomach, were 10.7 at the end of expiration, and 3.5 at the end of inspiration.

As the peristaltic contraction approaches the lower esophagus, the deglutition pressure complex changes slightly as it reaches the slightly dilated or bulbar portion of the terminal part of the body of the esophagus known as the ampulla. The complex is more prolonged due to the slowing of the rate of propagation of the peristaltic contraction, and the large positive pressure wave terminating the complex tends to decrease in amplitude.

With deglutition, the high basal pressure that exists in the esophagogastric junction decreases to form a slow negative wave of 7 to 10 sec. duration. This decrease in pressure occurs while pressure is increasing in the lower esophagus, thus greatly extending the pressure gradient from the lower esophagus to the esophagogastric junction. The slow decline in pressure is followed by an equally slow increase to form a smooth negative pressure wave. This is continued into a similar slow positive pressure wave which appears to occur in sequence with the peristaltic pressure wave of the lower esophagus. The high pressure zone of the esophagogastric junction extends slightly above and below the diaphragmatic hiatus, and pressure changes due to deglutition are similar in each part but are of less magnitude when recorded from the segment below the diaphragm than from that part located just above the diaphragm.

The lower esophageal sphincter thus relaxes reflexly, as does the cricopharyngeus muscle at the pharyngoesophageal junction, allowing esophageal contents to be emptied into the sphincteric area by the progressing high pressure gradient between the lower esophagus and the sphincter segment. The sphincter then contracts slowly with higher pressures in the part of the segment above the diaphragm than in that below. This provides for emptying of the sphincteric segment into the stomach. The motor function of the lower esophageal sphincter is distinct and separate from that of the body of the esophagus, however, the relaxation and contraction of this sphincteric segment in sequence with the advancing peristaltic contraction permits esophageal emptying into the stomach.

It has been demonstrated recently that the lower esophageal sphincter is under the humoral control of both gastrin and secretin. When small amounts of gastrin are administered intravenously or when the antrum is alkalinized to release endogenous gastrin, the lower esophageal sphincter markedly increases its tone. Secretin has the opposite effect. Pure intravenous secretin or acidification of the duodenum to release endogenous secretin causes a significant decrease in intrasphincteric resting pressure. When gastrin is released to produce high rates of HCl secretion, its effects on the lower esophageal sphincter thus helps prevent reflux of the acid contents of the stomach into the esophagus.

Disturbances of Esophageal Motility. Dis

turbances of esophageal motility usually result in dysphagia, or difficulty in swallowing, which may, or may not, be associated with pain Some of the more common causes of dysphagia will be discussed briefly

Diffuse spasm of the esophagus produces dysphagia which characteristically occurs intermittently and is associated with pain In this disorder, the esophageal musculature appears to be hypersensitive to the stimuli produced by deglutition and also by local distention Following deglutition, long segments of the esophagus may go into a marked spasm, which produces severe substernal pain When esophageal intraluminal pressures are measured in these patients, deglutition may result in a marked distortion of the normal pattern Intraluminal pressure may exceed 300 cm of water pressure

Achalasia results in a progressive dysphagia from an inability of the lower esophageal sphincter to relax reflexly in response to deglutition In this disorder, the primary peristaltic contractions originate in the pharynx in a normal manner, but when they have reached a level of the upper esophagus, they become ineffective, weak, and finally fade away Secondary contractions often occur due to the local distention by retained foods and secretions, but once initiated, these contractions move purposelessly up and down the esophagus for a short distance, then they too disappear Without normal propagation of peristaltic contractions, little intraluminal pressure gradient is created and esophageal transport is markedly retarded As the normal motor function progressively fails, the retention of solid foods, liquids, and secretions occur, and this leads to marked dilatation and elongation of the esophagus

Numerous investigators have demonstrated a marked degeneration of Auerbach's plexus, especially in the lower parts of the esophagus in patients with achalasia The cause of the degeneration of the myenteric plexus is unknown The unusual sensitivity of the esophagus in patients with achalasia to mecholyl supports the pathological findings that achalasia is due to a disorder of the intrinsic nerve plexuses When mecholyl is administered subcutaneously in small doses of only 1 or 2 mg , the esophagus responds by going into marked spasm (Kramer and Ingelfinger, 1951, Hightower, et al , 1954) The response to mecholyl has been used as a diagnostic test in establishing the diagnosis of achalasia

Scleroderma is an example of a systemic disease that may produce dysphagia by interference with the third stage of deglutition The primary disorder of esophageal motility is the defective propagation of the primary peristaltic contraction due to a myogenic failure Also, there is rather marked loss of muscular tone in the esophageal wall As in achalasia, scleroderma involves most of the esophagus, and is not confined to any particular segment, although changes are usually more evident in the lower esophagus

Few deglutition pressure complexes reach the lower esophagus and those that do are usually at low amplitude The esophagogastric junction is also involved in this disease, with a loss of the normal high pressure zone With the loss of the pressure barrier between the esophagus and the stomach, gastroesophageal reflux is a common finding Mecholyl does not increase esophageal intraluminal pressure in patients with scleroderma, as is observed in patients with achalasia

Hiatus hernia is a very common disorder that may produce dysphagia due to the anatomical alterations of the normal esophagogastric junction In these patients, the esophagogastric junction is displaced upward out of the diaphragmatic hiatus into the chest When the esophagogastric junction is displaced into the chest, the high pressure zone is decreased and reflux from the herniated portion of the stomach above the diaphragm into the esophagus is a common finding Some degree of peptic esophagitis is not at all uncommon and this probably accounts for the increased irritability and spontaneous contractions frequently observed in pressure records of the esophagus in patients with hiatal hernia

Chalasia is a term applied to a relaxed esophagogastric junction, the esophagogastric junction is not displaced out of the diaphragmatic hiatus as in hiatal hernia, but there is a decreased high pressure zone within the lower esophageal sphincter This permits frequent gastroesophageal reflux This condition is usually seen in infants but it may occur in adults

Obstruction of the esophagus due to *neoplasm* or *stricture* is a common cause of dysphagia In the case of neoplasms, dysphagia develops slowly but becomes progressively worse Difficulty is usually first noticed with solid foods, and later with liquids The patient will complain that foods "stick" or "hang-up" in the esophagus, and usually he can point to a

site that indicates the level of the lesion Strictures of the esophagus in children are frequently due to drinking caustic solutions

HEARTBURN AND BELCHING

Heartburn is a painful sensation located retrosternally in the midline between the xiphoid process and the manubrium Its duration is variable and it may radiate to the upper abdomen the anterior chest wall, the jaws and even into the arms Heartburn is frequently experienced by patients with hiatal hernia and occurs during the second half of pregnancy in approximately 50% of cases

Tuttle, Bettarello, and Grossman (1961) measured intra-esophageal pH in patients who complained of heartburn not associated with belching, and found that the heartburn was associated with a drop in pH to below 4 0 in the lower esophagus There was no evidence of disturbed esophageal motility when heartburn was induced by instilling hydrochloric acid into the stomach Siegel and Hendrix (1961), on the other hand, found pain and abnormal esophageal motility to occur at the same time when they perfused acid into the esophagus of patients with esophagitis The cause of "heartburn" is still not completely understood In patients with esophagitis, the instillation of a number of noxious substances into the esophagus may cause the distress Abnormal esophageal motility may or may not be associated with heartburn

Belching The tendency, after a meal, for small amounts of gas to be expelled from the stomach into the esophagus and mouth is experienced by most normal persons The gas in these instances is, as a rule, simply air which has been previously swallowed and it has the composition of atmospheric air Aerophagia is seen in the nervous type of subject, or in one who has some gastric discomfort, he resorts to the air swallowing trick in an effort to gain relief

In a study of the mechanism of belching (McNally, Kelley, and Ingelfinger, 1964), radiological and pressure recordings were made during the distention of the human stomach with air Pressures were recorded from both the esophagus and the stomach As the stomach was slowly distended with air there was an initial increase in intragastric pressure of 2 to 6 mm of Hg., but no further increase occurred, even when up to 1600 ml of air was introduced Within 3 min after the stomach had been distended with air they observed that sudden equalizations of gastric and esophageal pressures took place at inconstant and intermittent intervals in their subjects Cineradiography showed that the sudden equalization of gastric and esophageal pressures was associated with escape of air from the stomach into the esophagus Usually within 6 to 10 sec , a secondary esophageal peristaltic wave occurred, returned the refluxed air to the stomach, and reestablished the esophagogastric barrier The authors referred to this as "simple reflux"

Pressure changes associated with belching were different from those observed during simple reflux When belching occurred, a simultaneous pressure spike was recorded from the stomach and the esophagus in all subjects The pressure spike and the associated belching were superimposed on the pattern of simple reflux After the belch, the intraesophageal pressure remained elevated and identical with intragastric pressure for 6 to 10 sec , until secondary esophageal peristalsis emptied into the esophagus and returned esophageal and gastric pressure to their control level

Gastric Motility

Before attempting to describe the movements of the stomach, it is desirable to define the anatomical terms that will be used (Fig 2 51) The stomach consists of a *fundus*, a *body*, and a *pyloric portion* (antrum) The fundus is that portion of the stomach which, in the human in the upright position, lies above a horizontal plane passing through the esophageal opening The body of the stomach (corpus ventriculi) extends from the fundus to the *incisura angularis* on the lesser curvature, and to a corresponding notch on the greater curvature The incisura is a conspicuous notch on the lesser curvature, which serves not only to separate the anatomical divisions of the stomach, but also marks the position of a band of smooth muscle which has a higher degree of tonus than most of the remainder of the gastric muscle and causes a narrowing of the gastric lumen at this point The pyloric portion (pars pylorica) extends from the incisura angularis in the lesser curvature, and the corresponding notch on the greater curvature, to the junction of the stomach with the duodenum The pyloric portion differs from the remainder of the stomach in having much heavier musculature

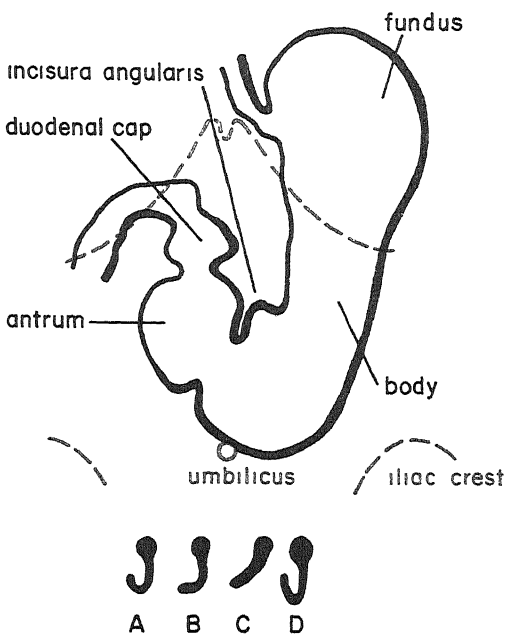


Fig 2 51 Diagram showing the position and subdivisions of the stomach *A B* and *C* represent *J* reversed *L* and steer horn types of stomach respectively *D* stomach of the J-shaped type with the greater curvature lying well below the level of the umbilicus Antrum is also known as the vestibule

The cardia is simply the esophageal orifice, that is, the opening between the esophagus and the stomach Properly speaking, it refers to no anatomical structure, nevertheless, the term is often used to refer to the entire region of the stomach surrounding the cardia, and in that sense it includes the mucous membrane and muscle in the immediate vicinity of the orifice

The pylorus is the opening between the stomach and duodenum and, like the cardia, refers to no anatomical structure However, the term is often used to designate the pyloric portion of the stomach Likewise, the pyloric sphincter is sometimes referred to as the pylorus These usages are incorrect and often give rise to misunderstandings The lumen of the pyloric portion may be divided into an antrum and a pyloric canal The separation between these two portions is marked by a sulcus intermedius when this is present The portion to the right of the sulcus is known as the pyloric canal, and the portion to the left as the pyloric antrum The antrum is not a structure but a space, nevertheless, the entire pyloric portion is often called the antrum in clinical literature

THE PERIODIC ACTIVITY OF THE EMPTY STOMACH

The smooth muscle of the stomach is seldom completely inactive, but its activity increases during gastric digestion and periodically when the stomach is empty

Types of Contractions Carlson (1913) recognized three types of contractions in the empty stomach

Type I occurs when the tone of the stomach is feeble It consists of single contractions superimposed on a low tone level The tonus of the stomach does not rise between contractions and there is always an interval of time between the end of one contraction and the beginning of the next Each contraction lasts for about 20 sec and the interval between contractions is from ½ sec to 3 or 4 sec When recorded by means of a small balloon in the antrum, type I contractions produce a pressure change of approximately 5 cm of water (Hightower and Code, 1950) They occur in both rhythmic and nonrhythmic patterns When present rhythmically and recorded from the antrum, their rate is constant at 3/min Carlson (1912) referred to this as the "20-second rhythm " It is felt that the type I waves of the stomach represent a "mixing" type of motor activity One gets the impression from balloon-kymographic records, and also from fluoroscopic observations, that this type of motor activity represents a kneading of the gastric contents

Type II contractions occur when the tonus of the stomach is strong Cannon (1911a) referred to these as "hunger" contractions There may be no intervening pause between the contractions and each lasts for 20 sec Type II contractions may occur also in a nonrhythmic pattern, and when they do, their duration is more variable, ranging from 12 to 25 sec Pressure changes produced by these contractions usually range from 10 to 50 cm of water, although at times pressures of 100 cm of water are observed Type II gastric contractions represent peristalsis In the antrum, this type of motor activity creates the pressure gradient that effects gastric evacuation into the duodenum

Type III gastric contractions have been described as an "incomplete tetanus " They are usually observed at the end of a period of gastric motor activity and are followed by a period of quiescence They occur rarely in man and seldom make up more than 1% of the activity observed in motility records The type III con-

traction is recorded as an increase in basal pressure lasting 1 to 5 min with rhythmic type II contractions superimposed upon the increase in basal pressure From a functional standpoint, it is thought that this type of motor activity represents a change in tonus or decrease in diameter of the lumen The increase in the basal pressure observed during a type III contraction is usually less than 10 to 15 cm of water

When recorded from man, after an overnight fast, antral gastric motility records show that some type of activity is present approximately 40% of the time and this activity is broken down as follows type I contractions about 25%, type II about 15%, and type III 1% during the period of observation Examples of the three types of gastric contractions are illustrated in Fig 2 52

DIGESTIVE PERISTALSIS

As soon as food is taken, the "hunger" contractions cease, and after a variable period of time, they are replaced by digestive peristalsis The outstanding characteristic of this type of gastric peristalsis is the constancy of its rhythm When the stomach contains food, peristaltic waves occur regularly, at a frequency of about 3/min in the human, 4/min in the dog, and about 5/min in the cat Each wave begins as a slight circular constriction of some part of the body of the stomach The exact point of origin varies with conditions, one of which is the amount of food in the stomach The contraction appears in the x-ray silhouette as a slight indentation on the greater and lesser curvatures which travels toward the pylorus, becoming deeper meanwhile, until it finally ends with a contraction of the pyloric sphincter

A peristaltic wave takes 1 min , more or less, to travel from its point of origin to the pylorus Since there are three waves per minute, the stomach will usually be occupied by three waves at one time, or by four or more if the origin is near the fundus (Fig 2 53) If the origin is low, only one or two may be seen

During digestion, the body and fundus of the stomach, which serve the reservoir function primarily, behave quite differently from the pyloric portion (which serves mainly the function of mechanical digestion and propulsion) The muscle of the body and fundus are relatively thin and participate to only a moderate extent in visible peristalsis It is this portion of the stomach that undergoes the greatest changes in size and shape, accommodating itself through changes in tonus to the volume of its contents

ELECTRICAL ACTIVITY OF THE STOMACH

When electrodes are placed on the muscle layers of the wall of the stomach, a rhythmic, cyclical, electrical potential charge can be recorded When multiple electrodes are placed in the long axis of the stomach it can be determined that the electrical activity is propagated from the fundus to the pylorus This cyclical potential charge is due to periodic depolarization of the cells making up the muscle layers in the wall of the stomach and has been termed "basic electrical rhythm" (BER) or "pace setting potential" (PSP) (Code and Carlson, 1968) The exact site of the "pacemaker" is not known but is presumed to be in the fundus near the cardia The basic electrical rhythm of the stomach is omnipresent, whether or not contractions of the wall of the stomach are present The cyclical rate of the potential change is exactly the same as the maximal rate of type I and type II contractions, i e , 3/min in the antrum of man When type I or type II contractions occur, a spike or burst of fast electrical activity occurs which is superimposed upon a cycle of BER In dogs, the rate of conduction of the BER at the junction of the antrum and body of the stomach is about 0 5

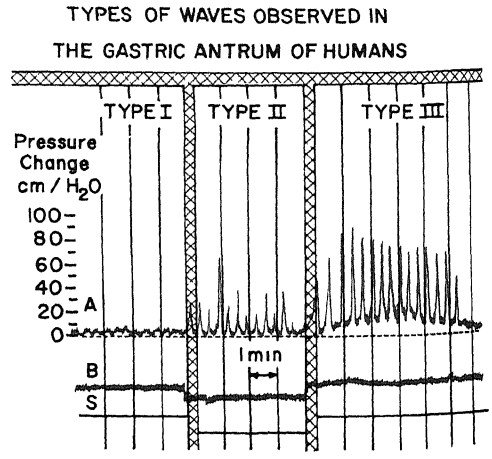


Fig 2 52 Types of antral motility recorded from normal human being Note that the rhythmic rate of type I and type II waves is identical at 3/min The type III complex consists of a rise in basal pressure on which is superimposed rhythmic type II waves (From Code et al 1952)

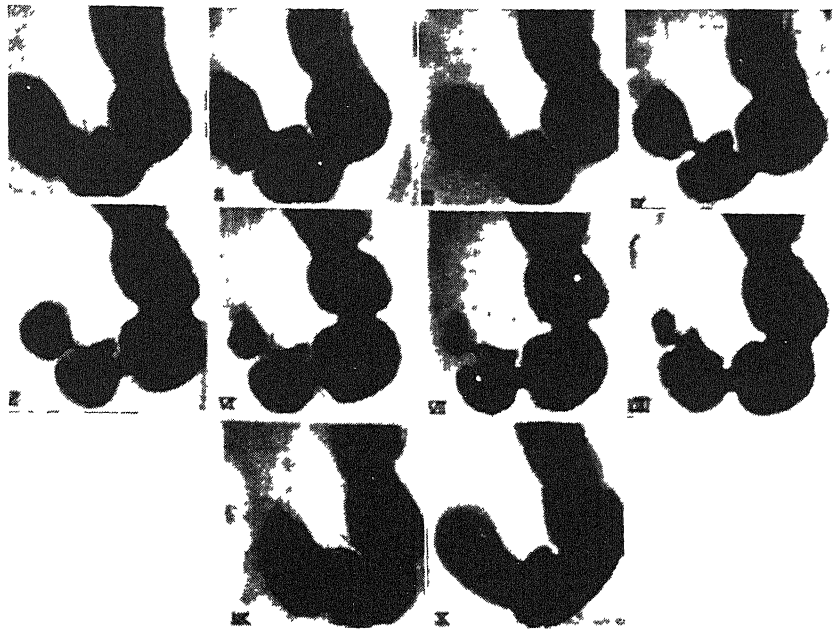


Fig 2 53 Serial x-ray photographs of human stomach taken at 2-sec intervals during digestion (From Cole 1917)

cm /sec , while in the distal antrum, the rate is 2 to 4 cm /sec Thus, conduction is increased as the pylorus is approached

THE GASTRODUODENAL PRESSURE CYCLE

When a peristaltic wave approaches the pylorus, the antral pressure increases, usually to between 15 and 30 cm of water, followed by an increase in bulbar pressure However, the bulbar pressure rises more rapidly than the antral pressure, so that the two reach their peaks simultaneously, after which they both fall again to basal levels This cycle is repeated, with only quantitative variations, with each gastric cycle Contraction of the pyloric sphincter usually begins near the start of the rise in bulbar pressure but outlasts the increase in bulbar pressure

In fluoroscopic studies, passage of gastric contents through the pylorus could be seen to begin as the antral wave approached the pylorus but before the elevation in antral pressure could be detected Evacuation continued through the early part of the rise in antral pressure but ceased some time during the succeeding rise in bulbar pressure (and the sphincteric contraction) Consequently, the time during which gastric contents could be seen passing into the duodenum could be divided into two distinct but continuous phases, one, during which only basal pressures served as a driving force and resistance at the pylorus was low (designated evacuation period "A"), and another during which the antral pressure substantially exceeded the duodenal pressure and the pyloric resistance was rising (called evacuation period "B" (Quigley, 1943, Quigley and Brody, 1950) (See Fig 2 54)

GASTRIC FILLING

It is possible to determine how solid or semisolid food arranges itself within the stomach of an animal by marking successive portions of the food with dyes or pigments so that they can be identified later After the stomach is filled, the animal is killed, and the stomach removed and frozen, or otherwise fixed, and sectioned In this way it has been found that the first food eaten lies against the greater curvature in the body and fundus of the stomach Successive layers lie progressively nearer the lesser curvature until the last portion eaten is near the upper end of the lesser curvature in the vicinity of the cardiac orifice Liquids, on the other hand, tend to remain near the lesser curvature of the stomach and to flow toward the pylorus along a V-shaped groove formed by the muscle in this region This is the so-called *Magenstrasse* which is evident in some animals, but less so in the human In any case, if a large

volume of liquid is swallowed, it flows around the food mass and tends to be distributed over the entire interior of the stomach between the food mass and the stomach wall

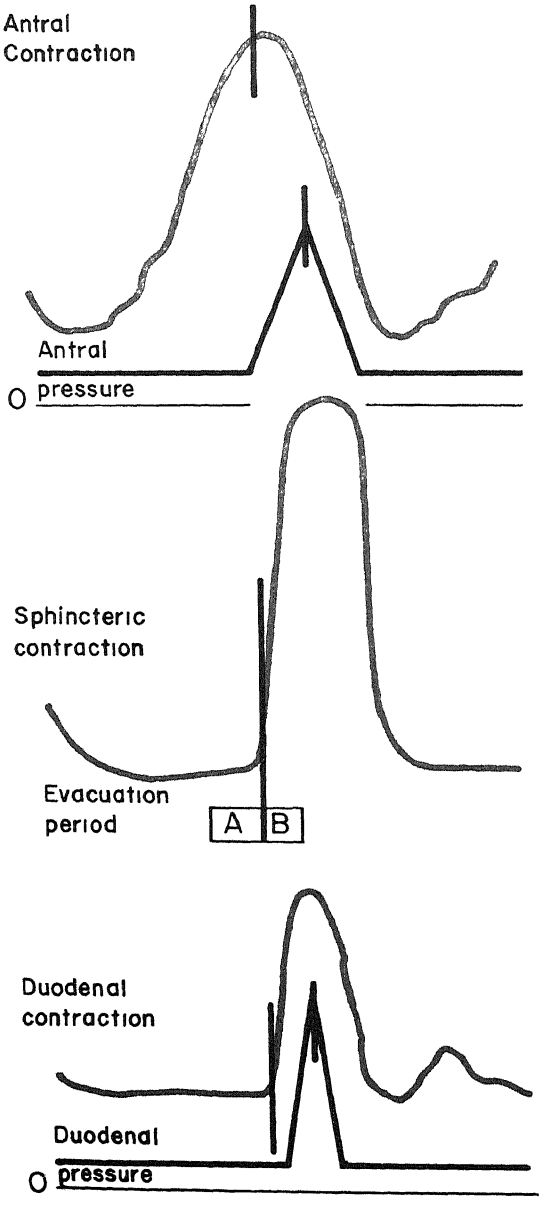


Fig 2 54 Composite diagram showing the successive contractions of the pars pylorica (antrum ') pyloric sphincter and duodenum during a single gastric cycle and the associated intraluminal pressure changes Evacuation periods A and B are also shown The contraction curves were traced from a kymographic balloon water-manometer record The pressure curves are approximate copies of the curves published by Quigley and Brody (1950) Vertical lines mark simultaneous points

CONTROL OF GASTRIC EMPTYING

A relatively small part of the energy expended by the gastric muscle is utilized in propelling food from the stomach into the intestine through the pyloric orifice The vigor of the gastric muscular contraction is such that if it were all utilized in propulsion the stomach would empty itself within a very few minutes By far the greater part of the work is utilized in mechanical digestion, serving to macerate the food particles and break them up into small fragments This work is done largely by the pyloric portion, which in some ways resembles in function the gizzard of birds The powerful contractions of the pyloric muscle drive the food against the pylorus All but the small part that escapes into the duodenum is forced back through a narrow orifice, formed by the contraction wave, into the body of the stomach It is this manipulation of the food mass which results in its being broken down and mechanically digested in the pyloric portion of the stomach

The rate at which food leaves the stomach is determined by a variety of factors, among which are the total volume of the gastric contents, and the consistency, chemical composition, pH, and osmolar concentration of the food Hunt and Spurrell (1951) have shown that for any individual, and for any type of meal, the amount evacuated in a unit of time is a function of the volume of the gastric contents The effect of consistency on the rate of emptying is best seen by comparing the emptying of liquids with the emptying of solids Inert liquids, which neither stimulate the stomach chemically nor osmotically, leave the stomach rapidly Water, for example, begins to leave the stomach almost as soon as it is swallowed Solids, on the other hand, are not normally evacuated from the stomach until they are reduced to a fluid or semifluid consistency Undigested solids are not normally allowed to pass the pyloric sphincter but are returned to the body of the stomach (Cannon, 1911b)

The influence of the chemical composition of the food on emptying is shown by the relative rate of emptying of the three major foodstuffs Cannon (1904) showed that for meals of a given size, a meal consisting mainly of carbohydrate left the stomach more rapidly than one consisting mainly of protein or fat, and that a protein meal left more rapidly than a fat meal Fat, especially if eaten in large quantities, re-

mains in the stomach for a very long time It is not unusual to find food in a dog's stomach as long as 24 hr after a meal consisting largely of fat meat

The osmolar concentration of the gastric contents has a profound effect upon the rate of gastric emptying One of the earliest facts to be determined with respect to gastric emptying was that physiological saline leaves the stomach faster than does salt solutions of either greater or lesser concentration Hunt (1956) has shown that in human subjects, adding salt to water hastened emptying of the water from the stomach, up to a concentration of about 100 mEq /liter Above 150 mEq /liter the emptying was delayed In this way, he showed that making the gastric contents either hypertonic or hypotonic with respect to the blood caused a delay in emptying This is an important factor in the slow emptying of sugar solutions although, as will be pointed out later, sugar slows gastric emptying in another way

The effect of acid on gastric emptying has been the subject of considerable study and extensive speculation It is well known that strong acid solutions leave the stomach slowly, and that the rate of emptying of such solutions can be accelerated by neutralizing the acid The acidity of the gastric and intestinal contents is usually not great enough during digestion to affect gastric emptying materially For example, a pH of 3 0 in the duodenum is necessary in order to influence the rate of emptying of the stomach, but the intestinal contents usually have a pH that is well above this level, pH values of 4 0 to 6 0 being most common Nevertheless, the inhibitory effect of acid on gastric emptying has an important function in protecting the duodenal mucous membrane against excess acidity

Enterogastrone. The inhibitory effect of fat on gastric motility has been known for many years and was recognized by Cannon Fat exerts its effect in the absence of any available nerve supply (Farrell and Ivy, 1926) Kosaka and Lim (1930, 1933), and Quigley, Zettelman, and Ivy (1934) showed that fat inhibition is caused by a hormone which Lim (1933) called *enterogastrone* This hormone is released from the intestinal mucous membrane by fat or fatty acids and some other substances in the duodenum While circulating in the blood, it is carried to the stomach, where it has an inhibitory effect on the gastric muscle

In summary, we can visualize the events following a meal approximately as follows gastric emptying begins as soon as any part of the gastric contents become fluid enough to pass through the pylorus Once started, it proceeds rhythmically a small amount being evacuated about every 20 sec in man Regulation of emptying begins as soon as the evacuated material has accumulated in the intestine to the point where any one of the numerous stimuli to the intestine associated with the chyme reaches threshold value These stimuli are comprised in the chemical constituents of the chyme and to some extent in its volume The chemical stimuli include the following, listed in the order of inhibitory potency in the concentrations in which they are usually encountered fats, fatty acids, proteoses, peptones, amino acids, sugars and other products of starch digestion, and hydrogen ions, also effective are osmotically active substances and nonspecific irritants The effect of these stimuli is to decrease the tone and peristaltic activity of the stomach, thus reducing the pressure gradient which develops with each gastric cycle and which is responsible for the passage of fluid through the pylorus With the more powerful inhibitors e g , fat, the gradient may be reversed, and duodenal contents will then be regurgitated into the stomach The inhibitory effect is exerted through a reflex involving the vagus nerves (*the enterogastric reflex*) or through a hormone (*enterogastrone*) released into the blood from the intestinal mucous membrane, or both The pyloric sphincter plays a part by preventing regurgitation of the duodenal contents into the stomach, it also contracts rhythmically in such a way that it apparently limits the volume evacuated at each cycle, but as yet there is no evidence that this function is utilized in regulating the overall rate of gastric emptying

ABNORMAL GASTRIC MOTILITY

Cardiospasm and *pylorospasm* are conditions in which these sphincter muscles are supposed to be in a state of sustained contraction Doubtless such abnormalities can occur but probably much less frequently than they are diagnosed Cardiospasm, for example, could be easily confused with loss of power in the esophageal muscle or absence of esophageal peristalsis, likewise, pylorospasm is often diagnosed when the true condition is failure of the gastric peristaltic mechanism to develop adequate gastric pressure, e g , following the operation of

vagotomy True pylorospasm occurs as a congenital condition in infants who suffer from hypertrophy of the pyloric muscle More accurately, the term pyloric stenosis or hypertrophic pyloric stenosis, should be applied to this condition Pylorospasm rarely occurs in adults but when present it is associated with irritative lesions in the region of the pylorus, either of the duodenum or pyloric portion of the stomach

VOMITING

Vomiting is a reflex which serves to relieve the upper gastrointestinal tract of its contents This may occur either because the contents are irritating, or because the organs themselves are more irritable than normally In either case, there is an unusual degree of stimulation of some part of the tract The most sensitive area is the first portion of the duodenum (Luckhardt, Phillips, and Carlson, 1919), but adequate stimulation of other parts of the intestine, or of the stomach, can induce the reflex Even mild mechanical stimulation of the pharynx or fauces is effective in most people Inflammation of mechanical disturbances in the intestine (appendicitis, obstruction, or strangulation) may induce violent emesis Vomiting may be induced by abnormal stimulation of sensory receptors in organs outside the gastrointestinal tract, such as the uterus, kidneys, heart, or semicircular canals Effective stimuli may also enter by way of the eyes, nose or mouth in the form of nauseating sights, odors or tastes, these give rise to what is called "psychic" vomiting It is not unusual for a soldier to vomit on going into battle, or at some time during the battle experience, some people vomit at the sight of blood

Vomiting that is induced by irritation of organs other than those of the gastrointestinal tract is sometimes referred to as "reflex" vomiting This is an incorrect distinction however, since nearly all vomiting is reflex in origin Even psychic vomiting is probably the result of conditioned reflexes

Vomiting may be induced by certain types of motion such as seasickness, air sickness and car sickness, or the nausea that some people experience when in a swing This is probably the result of unusual stimulation of the equilibratory sense organs of the labyrinth.

The Vomiting Center. The vomiting center lies in the dorsal part of the lateral reticular formation of the medulla oblongata close to (and to some extent including) the tractus solitarius It lies in close relation to (but is quite distinct from) the respiratory centers It is one of a constellation of visceral centers in this situation which includes salivary nuclei defecation, and vasomotor centers as well as the vestibular nuclei These form a relationship consistent with the physiological reactions (salivation vasomotor respiratory etc) associated with the vomiting act The *efferent* fibers are contained chiefly in the phrenics the vagi and the sympathetics but fibers are also conveyed by spinal nerves to the abdominal muscles and by cranial nerves to the muscles of the pharynx, palate, etc The *afferent* impulses reach the center along a multitude of routes the chief being the vagal and sympathetic fibers of the stomach and abdominal viscera

"Central" Vomiting Vomiting may occur as a result of direct stimulation of the vomiting center in the medulla Certain substances, such as *apomorphine*, *ipecac (or emetine)*, and *picrotoxine*, when administered intravenously, cause vomiting through a specific action upon the center These substances induce vomiting in animals when applied directly to the center Typical vomiting efforts (retching) follow the injection of apomorphine after removal of the stomach and intestines

The central vomiting caused by traumatic stimulation of the center due to head injuries, increased intracranial pressure, brain tumor, or meningeal irritation is most important clinically This type of vomiting is characterized by the great force with which the vomitus is ejected (*"projectile vomiting"*) There is also a minimum of nausea and a relative absence of participation of the voluntary muscles

The persistence of pernicious *vomiting of pregnancy* is probably due to increased excitability of the center resulting from some metabolic disturbance Ordinary morning sickness is looked upon as a minor form of this condition, and differing from pernicious vomiting only in degree Thus, morning sickness merges insensibly into the pernicious type of vomiting A neurotic element is also frequently a potent influence in the development and persistence of the condition The usual corrective measures are high carbohydrate feeding, the free administration of fluids, intravenous glucose if necessary, mild sedatives, and rest and quiet

The Vomiting Movements. Vomiting or

emesis usually begins with nausea, which is characterized by the secretion of excessive quantities of saliva containing a large amount of mucus, and certain well-known disagreeable sensations In the dog, it has been possible to analyze by graphic methods the sequence of events in vomiting The first muscular movement is a strong, sustained contraction of the upper small intestine Next, the pyloric sphincter contracts, then the pyloric portion of the stomach contracts These changes all take place during the period of nausea, and result in emptying the contents of the upper jejunum, duodenum, and pyloric portion of the stomach into the fundus and body of the stomach, which are relaxed and dilated After this has occurred, the voluntary muscles come into play, and at the same time the lower esophageal sphincter, esophagus, and upper esophageal sphincter relax Following an inspiratory movement, the glottis is closed and the abdominal muscles contract in a characteristic jerky manner (retching), compressing the stomach between the contracted diaphragm and the abdominal organs The resulting pressure on the gastric contents causes their evacuation through the relaxed esophagus

During the ejection of the vomitus the esophagus is relaxed throughout, the glottis is closed, the respirations are inhibited, and the larynx and hyoid bone are drawn upward and forward and are held rigidly in this position The throat is thus enlarged to allow free exit for the stomach contents, which are prevented from entering the nasopharynx by the elevation of the soft palate

Small Bowel Motility

The major part of the smooth muscle of the small intestine is distributed in a thick, circular layer and a much thinner, longitudinal layer The circular layer is responsible for most of the visible intestinal movements The outstanding characteristic of the muscle is its *rhythmicity*, which is manifest under appropriate conditions by alternate contractions and relaxations at a remarkably regular frequency

RHYTHMIC SEGMENTING CONTRACTIONS

The more common type of rhythmic contractions occurs at regularly spaced intervals along a section of the intestine, dividing it into short segments, also, they may occur singly, in pairs, or in a variety of other arrangements The area involved in each segment may be less than 1 cm In the cat, Cannon (1902) saw a certain type of rhythmic contraction in which a section of intestine was divided into short segments by rings of contractions occurring at regularly spaced intervals When the contracted areas relaxed, the relaxed areas between them contracted, dividing each segment into halves The two halves of adjacent segments joined together to form a new segment (Fig 2 55) Cannon described these contractions as *segmenting contractions*, and that is the name most commonly applied to them at present

In a study of duodenal motility in man, using simultaneous cineradiographic and pressure measuring techniques, Friedman and his associates (1965) observed two types of segmenting contractions One type consisted of a contraction localized in a segment less than 2 cm in length which was most marked on the free or external border of the duodenum and was therefore *eccentric* in appearance These eccentric contractions were observed to occur in two or three areas simultaneously or successively and were not followed in any sequential fashion by contractions in adjacent segments These local eccentric contractions rarely succeeded in emptying the area of the duodenum completely and were not associated with any change in intraluminal pressure From a visual point of view, they appeared to have a mixing function

The other type of segmenting contraction observed in the duodenum was *concentric* and consisted of a local contraction involving a segment usually longer than 2 cm and relatively uniform circumferentially These concen-

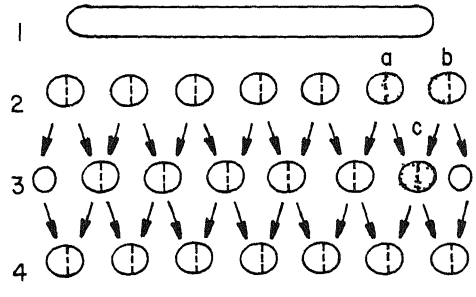


Fig 2 55 Diagram representing the process of rhythmic segmentation Lines *1 2 3 4* indicate the sequence of appearances in a single loop The dot lines represent the regions of division The arrows show the relation of the particles (e g *a* and *b*) to the segments they subsequently form (*c*) (After Cannon 1911b)

tric segmenting contractions were observed to empty the barium from the segment Most of the barium leaving the segment was displaced distally, although a small amount was occasionally displaced proximally As the segment relaxed, barium usually refilled the area, returning from both the proximal and distal adjacent segments The concentric segmental contraction was associated with an intraluminal pressure change that produced a positive monophasic pressure wave on the pressure record

The frequency of the rhythmic contractions varies with the species of animal and with the region of the intestine in which they occur For example, in the rabbit the contractions occur at a frequency of about 20/min in the duodenum and about 10/min in the lower ileum In the dog they occur at a frequency of about 18/min in the duodenum and 8 to 9/min in the lower ileum In man, segmenting contractions occur at a rate of 11/min in the duodenum, and in the ileum their rate is 8 to 9/min (Figs 2 56, 2 57) The frequency in other parts of the intestine is intermediate between these extremes, becoming less frequent as the distance from the pylorus increases The frequency is surprisingly constant in any one area and is not affected by stimulation of the extrinsic nerves or by neurotropic drugs In excised material the frequency varies with the temperature

The variation in frequency at different levels of the intestine suggested to Alvarez and his

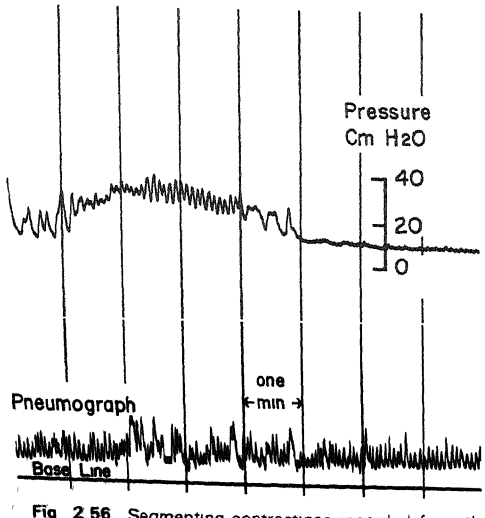


Fig 2 56 Segmenting contractions recorded from the duodenum of man by means of balloon-kymographic system Note that the rate of the rhythmic contractions (type I waves) was 11/min. (From Hightower 1959)

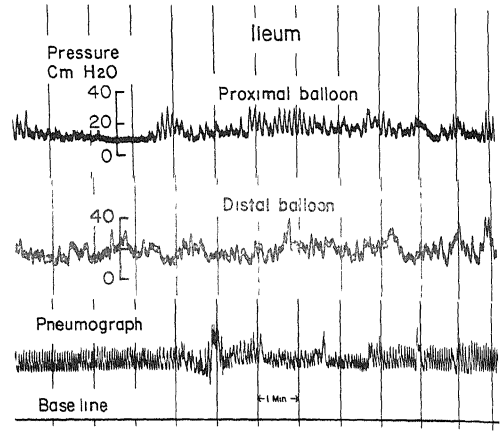


Fig 2 57 Motility tracing obtained from lower ileum in a patient with an iliac stoma A tandem system of balloons was used The proximal balloon record shows short bursts of rhythmic segmentation with the rate of contractions being 8/min (From Hightower 1959)

pupils the idea that there is a *gradient* in the various physiological properties of the intestine, activity being highest at the duodenal end and lowest at the ileal end of the small intestine Among the properties for which a gradient has been established are rhythmicity, irritability, shortness of latent period, susceptibility to drugs, and others of less obvious importance (Alvarez and Hosoi, 1929) The most easily observed is the gradient of rhythmicity manifested by the variation in frequency of rhythmic contractions (Fig 2 58)

The studies of Ambache (1947), Bozler (1949), and Milton and Smith (1956) have described the electrical activity of the intestine as consisting of slow waves which are frequently, but not necessarily, associated with local muscular contractions, and rapid, spikelike waves which appear only when the muscle contracts The slow waves have the same frequency as the rhythmic contractions, whether or not these are present, and (according to Milton and Smith) they are regularly conducted along the intestine in an aboral direction

The electrical activity of the small intestine is due to periodic depolarization of the muscle cells in the gut wall The BER is thought to originate in a "pacemaker" located near the ampulla of Vater in the duodenum The BER is propagated down the intestine in the longitudinal muscle layer The spike potentials occur when the circular muscle contracts and they are superimposed upon the slow wave activity

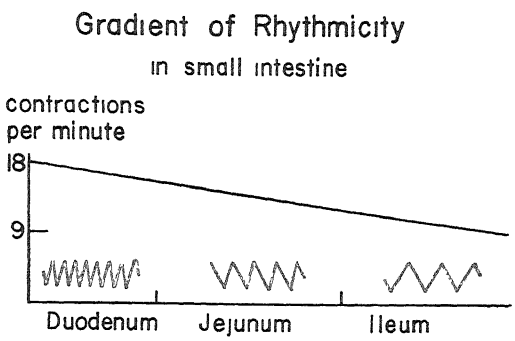


Fig 2 58 Diagrammatic representation of the gradient of rhythmicity in the small intestine The zig-zag lines are stylized representations of kymographic records of the contractions and relaxations of segments taken from the indicated levels of the intestine The sloping line suggests a linear distribution of the gradient however there is no direct evidence that this is so

Bass, Code, and Lambert (1961) have studied the electrical activity of exteriorized loops of duodenum in trained dogs Intraluminal pressures were recorded simultaneously to detect motor activity They found, like others, that cyclical changes in the electrical potential of the duodenum occurred This slow cyclical change was designated as the basic electrical rhythm (BER) In addition, irregular bursts of spike potentials were observed to be superimposed upon the BER When recorded from one site the cyclical changes in the BER were observed to be identical with the maximal rate of rhythmic type I contractions (18/min) When recorded from different levels of the duodenum the BER was observed to be propagated down the intestine at a rate of approximately 20 cm /sec The BER was recorded continuously, whether or not the intraluminal pressure catheter indicated the presence of motor activity

Spike potentials were noted in about one-third of the cycles of the BER They occurred in bursts of one to three spikes and were always located on the descending or negative slope of the slow potential change When spike potentials occurred, they were uniformly associated with an increased pressure within the lumen of the gut There was also an obvious correlation between the number and amplitude of the spikes and the magnitude of the increase in pressure in the gut, indicating a relationship between the number and amplitude of the spikes and the strength of the contractions occurring in the gut wall

PERISTALSIS IN THE SMALL INTESTINE

Stimulation of the intestinal smooth muscle causes contractions which tend to travel from the stimulated point in both directions However, under normal circumstances the progress of the contractions in an oral direction is quickly inhibited and the contractions disappear The contractions which travel in an aboral direction tend to persist, so that the intestine may be said to be polarized in the sense that conducted contractions travel normally in only the aboral direction Such aborally conducted contractions are commonly described as peristalsis The term has been used to describe a variety of contractile phenomena which have only one thing in common, namely the direction of travel The common conception of peristalsis as consisting of a wave of contraction progressing smoothly over otherwise relaxed intestine is erroneous Usually the peristaltic movement is superimposed upon the rhythmic segmenting contractions in such a way that the two are present simultaneously The peristaltic contraction makes itself manifest as a rise in the tone level of the intestinal muscle, without any interruption in the rhythm of the segmenting contractions, indeed, these are usually increased in amplitude at the peak of the wave This has the effect of narrowing the lumen of the intestine at the point at which the increase in tonus occurs and it may actually obliterate it As this wave of contraction, or tonus, travels down the intestine, it tends to sweep the contents forward, or toward the distal end of the intestine. In another type of propulsive movement, rhythmic contractions, although retaining their character, recur at successively more aboral levels and thus appear to travel along the intestine as a wave of peristalsis Ordinarily, peristaltic waves appear in the small intestine at irregular intervals and travel for varying distances (Bayliss and Starling, 1899), some travel only a few inches, others a few feet

In the duodenum of man, Friedman et al (1965) observed that some concentric segmenting contractions, instead of remaining localized in one area, would move distally over the duodenum into the jejunum When these peristaltic contractions occurred, pressure waves were recorded in a sequential manner, from above downward, by the three intestinal catheters localized in the duodenum The speed of the peristaltic contractions or stripping

waves were about 1 cm /sec Figure 2 59 illustrates intraluminal pressure changes occurring in the duodenum with peristalsis

The Law of the Intestine Bayliss and Starling (1899, 1901), to whose classical studies we owe most of our basic knowledge of intestinal motility, observed that the response of the small intestine to local stimuli consists of a contraction of the smooth muscle above and relaxation below the stimulated area "This," they said, "is the law of the intestine" Since the contraction wave and the preceding inhibition usually travel down the intestine in an aboral direction from the stimulated point, Bayliss and Starling thought that intestinal peristalsis might consist simply of a succession of responses to stimuli manifesting the law of the intestine In their opinion, a particular mass might give rise to stimulation resulting in contraction above the stimulated point, with relaxation below, and this would cause the mass to move aborally In its new position it would also act as a stimulus and again cause contraction of the muscle above and relaxation below, again being moved by this to a new, lower level A succession of such responses could according to them, constitute a peristaltic wave

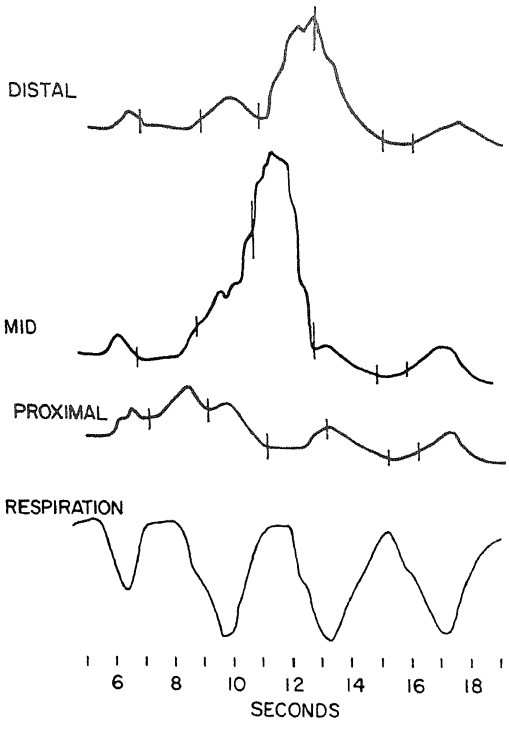


Fig 2 59 Duodenal intraluminal pressure changes with peristalsis The recording sites designated as proximal mid and distal are located at 2-cm intervals The short vertical lines indicate the frames of a cine sequence that were recorded simultaneously This series of three monophasic contractions occurring in timed sequence was associated with the transport of barium in an aboral direction In this patient the peristaltic contraction traveled at about 1 cm /sec (From Friedman et al 1965)

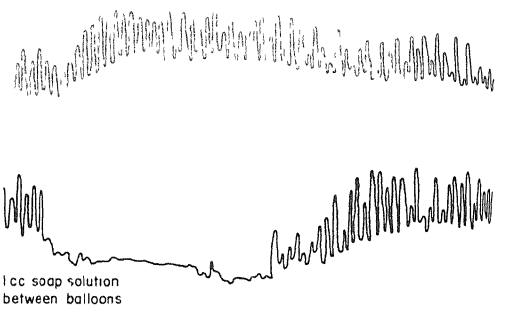


Fig 2 60 Myenteric reflex in a dog s duodenum Water manometer records from two balloons 5 cm apart in the lower duodenum Upper record is from the upper balloon A stimulus consisting of 2 ml of 2% soap solution in the intestine between the balloons caused contraction above (over the upper balloon) and relaxation below (over the lower balloon) the stimulated point (Thomas 1955)

Cannon (1912) proposed that the reflex responsible for the law of the intestine be designated the *myenteric reflex* (Fig 2 60) The myenteric reflex disappears if the intestine is paralyzed by applying such drugs as nicotine or cocaine to the serosal surface Presumably these drugs penetrate the longitudinal muscle layer and paralyze the underlying nervous mechanism Under these circumstances, the polarity of the intestine (Fig 2 61) is also abolished, and contractions initiated by local stimulation travel equally well in the oral and aboral directions Evidently, both the myenteric reflex and the polarity of the intestine are dependent upon the integrity of the myenteric plexus Of course, this does not prove that intestinal polarity is due to the myenteric reflex In fact, most observers have failed to find evidence of the occurrence of a myenteric reflex in ordinary peristalsis in the intestine Although the *Roll bewegungen* or rush waves are usually preceded by a phase of inhibition, the inhibitory phase has been difficult to demonstrate in spontaneous peristalsis of the intestine

THE INTESTINAL VILLI AND THE MUSCULARIS MUCOSAE

The absorbing surface of the small intestine is much greater than the serosal surface because

of the superabundance of the mucosa which under normal conditions, is thrown into circular folds called *plicae circulares* A still greater increase in surface is provided by the *intestinal villi* These are titlike projections of the mucous membrane about 1 mm in length which occur everywhere on the mucosal surface and which macroscopically resemble the pile on a piece of velvet Our interest is these structures at this point has to do with the activity of the *muscularis mucosae*, which lies everywhere beneath the mucous membrane, following each fold, and extending into each villus as a filament of smooth muscle fibers In the fasting animal, the villi are inactive and lie flat on the mucosal surface when first exposed for study Covering them with a liquid, such as a solution of sugar, salts, amino acids, weak alkalies, or even with distilled water, causes them to become active In the fed animal, activity of the villi can be observed without special stimulation The activity is of two kinds a lashing movement and a more or less rhythmical shortening and lengthening These movements accelerate the flow of blood and lymph and increase absorption The movements of the villi are partly under nervous control, being augmented by stimulation of the splanchnic but not the vagus nerves, and by sympathomimetric drugs They are depressed by atropine and nicotine Their activity is also increased by *villikinin*, one of the postulated intestinal hormones

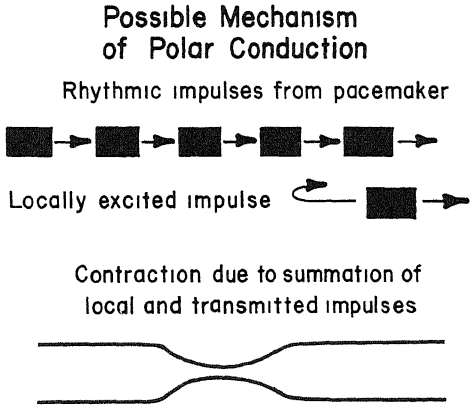


Fig 2 61 Mechanism of polar conduction in the intestine suggested by physiological evidence for a pacemaker mechanism in the upper intestine The block rectangles represent excitatory impulses those in the upper row presumably arose in the pacemaker and are being conducted analward A local excitory process cannot be conducted oralward because of the refractive phase following each conducted impulse from the pacemaker (Thomas 1955)

DISTURBANCES OF SMALL BOWEL MOTILITY

Disturbances of small bowel motility are rare, except when due to infections or inflammatory conditions, and in these instances there is increased motor activity of the propulsive type usually associated with diarrhea In some malabsorptive states, such as nontropical sprue small bowel transit time may be slow Abnormal motility patterns in the small intestine are seen in some nervous and endocrine disturbances and also in vitamin deficiency Their significance when they occur is not always clear Small bowel motility may be affected by intraluminal contents Saline cathartics markedly increase small bowel transit time due to the hyperosmolarity From a clinical standpoint, paralytic ileus and intestinal obstruction deserve further consideration

Paralytic Ileus. Although the small intestine usually shows some degree of activity in the intact animal, a dramatic and persistent quiet ensues when the abdomen is opened It is rare to see activity in the exposed intestine at surgery unless stimulating drugs are given Abdominal surgery, particularly if it involves handling the intestine, is apt to be followed by a period of *ileus*, as the loss of motility is termed clinically Peritoneal irritation from any cause, including bacterial peritonitis, may cause ileus As a rule, recovery is prompt following removal of the cause but if the condition persists for some time it tends to be self-perpetuating, due to the paralytic effect of the gaseous distention that usually develops When this occurs, removal of the gas through an intestinal tube of the type described by Miller and Abbott (1934) may save the patient's live (Fig 2 62)

The cause of the postoperative ileus is reflex inhibition of motility through reflexes from the irritated peritoneum in which the splanchnic nerves are an essential part of the reflex path Cutting the splanchnic nerves prevents the ileus

Intestinal Obstruction Symptoms of intense severity result when the lumen of the small bowel is completely obstructed as a result of constriction by an adhesive band, kinking, twisting or pressure by new growth, intussusception, strangulation by a hernial ring, etc The

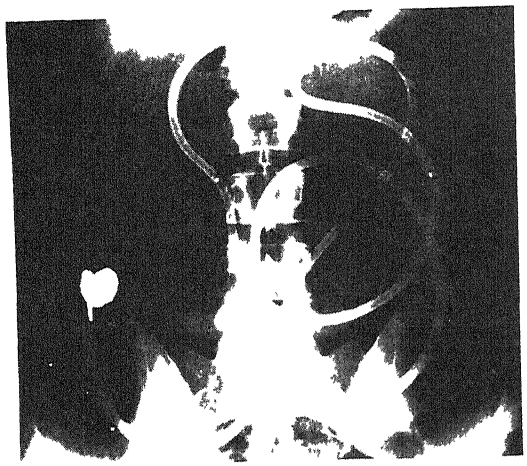


Fig 2 62 X-ray of abdomen showing Miller-Abbott tube in the small intestine

condition is ushered in by severe cramplike abdominal pain, vomiting, and shock If the condition is not relieved by operation, reverse peristalsis arises above the point of obstruction, intestinal contents pass into the stomach, and vomiting becomes fecal in character Later, the bowel above the obstruction loses its tone and becomes dilated and filled with intestinal secretions and gas The loss of fluid in the vomitus and the drainage of large quantities of fluid into the distended bowel leads to a *fall in blood chloride, alkalosis, and dehydration* The former is the result of tissue destruction combined with impairment of renal function Great prostration occurs, ending in death

Experimental obstruction When the intestine of an animal is tied across, the symptoms which follow are chiefly weakness, prostration, and vomiting The animal shows little or no evidence that it is suffering pain The higher in the intestinal tract the obstruction is made, the more severe are the symptoms, and the shorter is the duration of life after the operation Following obstruction of the colon the animal may survive for some weeks, whereas after obstruction of the jejunum or duodenum it dies as a rule within 5 or 6 days

The importance of distention as a factor in acute obstruction of the intestine is now generally recognized, and it has become the practice to decompress the bowel by suction-drainage through a tube passed from the nose to the small bowel and allow it to be carried to the level of the obstruction The tube has a double lumen and is provided at its tip with a small balloon which can be moderately inflated after it has entered the duodenum The balloon serves as a stimulus to motility which carries it and the tubing along the bowel to the point of obstruction (Fig 2 62)

THE ILEOCECAL VALVE

Most textbooks of anatomy describe the ileocecal valve as a slit-like orifice with a mucosal fold above and below on the colonic side of the orifice After death, the ileocecal junction does look like this when the cecum is opened at autopsy During life, however, if the cecum is opened during surgery, the ileocecal junction will be seen to be an invagination of the terminal ileum into the cecal wall, producing an appearance remarkably like, in both size and shape, the cervix of the uterus extending into the vagina The valve can withstand a high pressure in the cecum but yields to a low pressure from the ileal side Materials introduced by enema may in some instances pass through the valve into the lower ileum The ileocolic opening has been observed in man through a cecal fistula It appears as an oval or round opening from 2 to 3 cm in diameter situated in the center of a small papilla When tightly closed, the valve was found to offer considerable resistance to the passage of the finger While digestion was in progress, the papilla was observed to flush, its color altering from a pale pink to a vivid red The orifice opened rhythmically at frequent intervals and allowed a jet of fluid to escape into the cecum Emotional excitement or the swallowing of food increased the frequency of the ejections (*gastroileal reflex*) During fasting, nothing passed through for long periods, but in ½ to 4 min. after food was taken into the mouth, fluid appeared in gushes of about 15 ml every half minute or so

Kelley et al (1965) have made pressure studies of the ileocolonic junctional zone in dogs He prepared his animals by making ileal and colonic stomas and then measured pressures on either side of, and within, the ileocolonic junctional segment When a water-filled catheter tip, for recording pressures, was withdrawn at 0 5-cm intervals from the colon into the ileum of the fasted, conscious animal, a zone of high pressure was demonstrated at the ileocolonic junction. Mean pressures at the center of the junctional area were 20 cm of water above ambient The length of the high pressure zone was approximately 1 cm

The functions of the ileocolic valve appear to be 1) to prevent the contents of the ileum from passing into the cecum before the digestive processes have been completed, and 2) to serve as a barrier which prevents the bacteria-laden contents of the large bowel from contaminating the small intestine

Large Bowel Motility

The colon receives the mixed residues remaining after completion of intestinal digestion and absorption These consist of undigested or undigestible food residues, whatever remains of the digestive secretions including considerable quantities of water, and the fluid that has been swallowed or secreted and has escaped absorption in the small intestine The colon extracts the water and electrolytes from this mixture and forms the indigestible residues into fecal masses which are later evacuated The motor activities are appropriate to these functions They may be divided into two classes first, those that appear to be designed primarily to mix colonic contents and facilitate absorption, and second, those concerned principally with propulsion

COLONIC MOTILITY IN MAN

Four types of motility waves may be recorded with balloon or direct pressure measuring systems from the colon of man (Code et al , 1954, Hightower, 1959) When balloons are used to record colonic motility, they are usually water-filled, under low pressure, and small in size, with a length of approximately 5 cm and a diameter of approximately 3 cm Frequently, two or more such small balloons may be used in tandem to record colonic motility, being separated by 5 to 7 cm

Type I waves are small, simple, monophasic, positive pressure waves of low amplitude and short duration, similar to the same type of wave occurring in other parts of the alimentary canal The exact function of the type I wave is not known but presumably it serves as a component of motor activity which contributes to mixing of the colonic contents Type I waves usually have an amplitude of 5 to 10 cm of water pressure and their duration varies from 5 to 10 sec When occurring in a rhythmic pattern, the rate varies from about 8 to 12/min , depending upon the site from which they are recorded

Type II waves are similar to type I They are, however of much greater amplitude and of longer duration These waves are usually 25 to 30 sec in duration and the amplitude usually varies from 15 to 40 cm of water pressure This type of activity accounts for the major portion of total colonic motility They sometimes occur in rhythmical patterns with a rate of about 2 waves/min The type II waves are caused by the contractions in the large bowel which produce the haustra seen by radiologists Type II waves represent, predominantly, a mixing type of motor activity, although they may contribute to some extent to propulsive motility

Type III waves are of a complex pattern and consists of an increase in basal pressure or a change in tonus on which superimposed type I or type II waves, or both, appear These waves usually last from 1 to 4 min and the pressure change in base line tonus is usually less than 10 cm of water The exact function of the type III wave is not clear but it probably is predominantly mixing in its action

Type IV waves result from the "mass movement" activity of the colon These waves are rarely recorded from a normal individual unless the recording is obtained shortly after or during the act of eating These waves rarely occur in a rhythmical pattern, except in patients with chronic ulcerative colitis, and then the rate is usually about 1 wave every 2 to 3 min The amplitude of the type IV wave may reach 80 to 100 cm of water pressure The duration of the wave is characteristically prolonged and lasts from 2 to 4 min Their configuration is rather characteristic in that there is a fairly rapid increase in pressure at the beginning of the wave, the peak is reached, and then there is a rapid decline for 50 to 60% of the elevation The rate of decline changes and becomes much slower, producing a prolonged wave See Figures 2 63 to 2 66 for types of colonic motility waves recorded in man

The mass movements of the colon serve to empty the contents of the proximal colon rapidly into the more distal portions Frequently such movements are followed by a desire to defecate, and it is assumed that, on these occasions, the movement has been vigorous enough to carry fecal masses into the rectum Since normally a desire to defecate does not follow every mass movement of the colon, it must be assumed that, on some occasions, the forward movement of the colonic contents stops short of the rectum People differ considerably with re-

spect to their susceptibility to the gastrocolic or gastroileac reflex

In normal individuals who have fasted for 8 to 12 hr, some type of motor activity can be recorded from the colon for approximately 35 to 40% of the period of observation The predominant activity is the type II wave and accounts for 90 to 95% of the total activity Type I waves usually occur during 1 to 2% of the period of observation, and type III waves, 2 to 4% Type IV waves are rarely observed normally and are present for less than 1% of the period of observation

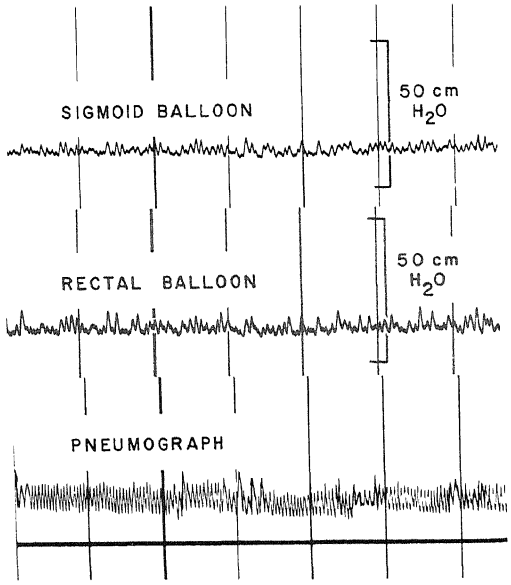


Fig 2 63 Type I waves Record of sigmoidal and rectal motility obtained from a normal individual Short bursts of rhythmical type I waves were observed with a rate of 11/min Distance between vertical lines is 1 min camera speed 0 4 mm /sec (From Hightower 1959)

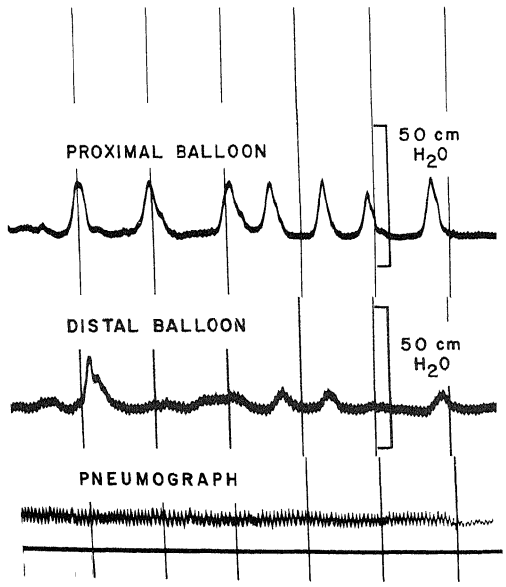


Fig 2 64 Type II waves Record obtained from transverse colon in a patient with a colonic stoma showing type II waves The balloons were separated by a distance of 7 cm Camera speed and time lines were the same as Figure 2 6 3 (From Hightower 1959)

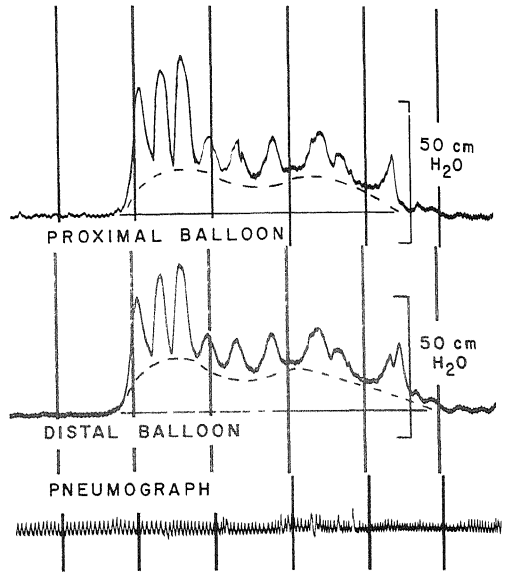


Fig 2 65 Type III waves Record from descending colon via colonic stoma Note rise in basal pressure with superimposed type II waves Both balloons were recording the type III complex indicating that a segment of colon at least 7 cm in length was involved in the motility being recorded Camera speed and time lines same as Figure 2 63 (From Hightower 1959)

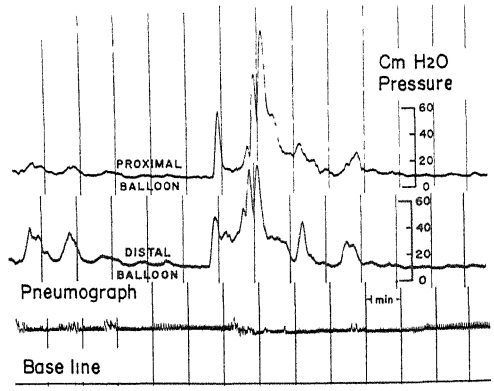


Fig 2 66 Type IV waves This type IV contraction was recorded from the sigmoid colon of a patient with chronic ulcerative colitis Note that pressure exceeded 100 cm water in the proximal balloon Type II waves are present at the left side of the tracings

Alterations of Colonic and Rectal Motility The effects of *emotions* on colonic motility has been extensively studied (Almy et al., 1947, 1949a, 1949b, 1950, Grace, Wolf and Wolff, 1948, 1951) Discussion of life situations which produce emotional conflict resulting in feelings of anger, anxiety, apprehension, resentment, hostility, or guilt cause increased motor activity However, when fear, fright, or dejection is the predominant feeling of the subject, hypomotility has been reported In patients with *chronic ulcerative colitis*, colonic motility records show rather characteristic differences from the normal There is a marked increase in the incidence of type IV waves, associated with propulsive action leading to the expulsion of fecal matter Type I activity is also increased There is a significant decrease in type II activity and usually a complete lack of type III waves Thus, total motor activity is decreased in these patients, due mainly to the reduction in the mixing type of motor activity At the same time, propulsive motility is increased

The Effects of Drugs on Colonic Motility. Many drugs influence the motor activity of the colon *Anticholinergic agents* of the quaternary amine type are potent inhibitors of colonic motility *Parasympathomimetic agents*, such as neostigmine and methancholine, stimulate colonic motility Neostigmine is particularly effective in producing a predominantly propulsive type of motor response *Morphine* increases the amount of nonpropulsive motility and tends to eliminate propulsive activity Morphine also markedly increases the tone of the large bowel Most of the so-called *antispasmodic and spasmolytic agents*, when administered by mouth in the recommended therapeutic doses, have little or no effect on motility of the large bowel

DEFECATION

The rectum is normally empty The fecal masses driven forward by the mass movements of the colon are stored, not in the rectum, but in the sigmoid or pelvic colon Fecal matter does not collect elsewhere if defecation is regular, however, if defecation is long deferred, the descending colon becomes filled when the pelvic colon can hold no more The desire to defecate occurs when, as a result of a mass movement, some feces enter the rectum An increase in rectal intraluminal pressure of 20 to 25 cm of water will usually elicit the desire to defecate While the usual stimulus is the taking of food, a glass of warm water, a cup of coffee, or smoking a cigar or cigarette may have the same effect Many people find that a glass of cold orange juice, or lemon juice and water is an effective stimulus to the gastrocolic reflex The desire to defecate may be induced by straining efforts, which may raise the abdominal pressure to as much as 200 mm of mercury and force feces into the rectum

When the rectum is filled with solids (feces), liquids, or gas to an extent which raises the intraluminal pressure to 20 or 25 cm of water, the desire to empty the rectum is experienced The receptors within the wall of the rectum are not only able to detect increases in pressure (pressorectors) but they can also, with a high degree of accuracy, differentiate whether the increase in pressure is due to feces, liquids, or gas The discriminating ability of these sensors is relied upon heavily by individuals who risk flatulating in a public place In patients with infectious diarrhea or proctitis due to ulcerative colitis or other inflammatory conditions, the ability to discern the contents of the rectum is frequently lost, and these patients find the risk is too great to relieve themselves of an increase in rectal pressure which is suspected of being due to gas

The act of defecation is, in the adult, preceded by a voluntary effort consisting of assumption of an appropriate posture, voluntary relaxation of the external anal sphincter, and usually, compression of the abdominal contents by means of straining efforts These movements in turn probably give rise to stimuli which augment the visceral reflexes, although these originate primarily in the distended rectum As a result of these reflexes, a mass contraction may involve the entire colon and the internal anal sphincter relaxes The mass contraction of the colon carries its contents into the pelvic colon, which in turn transfers them into the rectum, eventually to be evacuated by way of the anus, hence, the entire distal colon from the splenic flexure to the anus may be emptied at one time A prominent mechanical feature of the final act of evacuation is contraction of the longitudinal muscle of the distal colon, which is more pronounced in the rectum itself The shortening of the distal colon tends to elevate the pelvic colon and obliterate the angle which it normally makes with the rectum This straightening of the passage doubtless facili-

tates evacuation Shortening of the rectum itself is an important factor in expelling the feces through the anal orifice

The act of defecation provides another instance of a reflex that is under some degree of voluntary control The voluntary regulation consists of the ability to inhibit the reflex under normal circumstances and to initiate it voluntarily, provided the necessary visceral stimulus (recent distention of the rectum) is present Reflex centers for defecation have been located in the hypothalamus, in the lower lumbar and upper sacral segments of the spinal cord, and in the ganglionic plexus of the gut The intrinsic intestinal plexus of the lower colon is able to take over the function of controlling defecation, even in animals in which the spinal cord has been destroyed For this reason, in severe spinal injury the defecation reflex suffers less than does, for example, the reflex of micturition

The Composition of the Feces According to Cantarow and Schepartz (1954, p 276), adults on an average diet excrete from 75 to 170 g of feces daily, about 25 to 30% of which is made up of solids, the remainder is water If the diet is rich in vegetables the quantity is greater, especially if the vegetables are eaten raw Bacteria comprise about one-third of the dry weight of the feces under average conditions The pH is 7 0 to 7 5 The feces contain those inorganic substances which are relatively insoluble at alkaline pH ranges, such as calcium phosphate and oxalate, iron phosphate and similar compounds The organic material is principally cellulose, protein, and fats About half the protein nitrogen of the feces is of bacterial origin, the remainder represents unabsorbed intestinal secretions and digested fluids, mucus, and desquamated epithelial cells from the mucosa Only a small amount is actual food residue Enzymes are also present but in very small quantities, these are pancreatic amylase and trypsin, nucleases, maltase, sucrase, lipase and lysozyme Fats comprise 5 to 25% of fecal solids under normal circumstances They are greatly increased in conditions in which there is interference with pancreatic or biliary function In the total absence of pancreatic juice from the intestine, as much as 90% of the ingested fat may be excreted in the stools Much of the fecal fat has undergone digestion, as indicated by the following average composition of fecal lipids neutral fat 7 3% of dry weight of feces, free fatty acids, 5 6% and soaps 4 6%

The color of the feces is due chiefly to bile pigments, stercobilin (*urobilin*), which is produced by reduction of bilirubin by the intestinal bacteria The characteristic odor is due largely to indole and skatole, but is contributed to also by hydrogen sulfide and other odoriferous substances

Gas in the Intestine A variable amount of gas is always found in the large intestine Gas is seldom observed in the small intestine Intestinal gas consists of oxygen, carbon dioxide, nitrogen, methane, hydrogen, and small amounts of other gases including the rare atmospheric gases The average composition according to Pogrund and Steggerda (1948) is as follows carbon dioxide, 7 5%, oxygen, 3 0%, methane, hydrogen and other rare gases, 9 5%, and nitrogen, 80% The source of the oxygen, nitrogen, and carbon dioxide is swallowed air and diffusion from the blood Bacterial fermentation and putrefaction of food residues in the intestine produce the methane and hydrogen No matter what the composition of swallowed air may be, it quickly comes into equilibrium with the venous blood with respect to oxygen and carbon dioxide, and more slowly with respect to nitrogen, hence, the relative constancy of composition of intestinal gas with respect to these gases The amount of putrefactive gas is much more variable It is the rapid expansion of these gases that gives rise to intestinal cramps on ascending to high altitudes in an unpressurized plane Some of this gas is absorbed into the blood and some escapes from the rectum as flatus The amount in the intestine can be reduced by giving the subject pure oxygen to breathe

ABNORMAL COLONIC MOTILITY—CONSTIPATION

The most frequent manifestation of malfunction of the large intestine is constipation For a discussion of the various causes and the treatment of this condition, appropriate clinical texts should be consulted It may be mentioned here that one of the most frequent causes of chronic constipation is neglect of the call to stool The desire to defecate is experienced whenever the pressure within the rectum is increased An increase in the volume of the contents of the rectum of from 15 to 25 ml has

been shown to be sufficient to produce an urge. If the rectum is not emptied at such a time, the sensory mechanism becomes adapted to the pressure and the urge is no longer felt; it then requires further increase in volume with a corresponding increase in pressure to produce a sensation. This will also pass if neglected and a still further increase in volume is required to arouse a sensation; finally the irritability of the rectum is so reduced that an urge is no longer induced. Usually the subject will then take a cathartic which further injures the sensory mechanism. Frequent repetition of this cycle results ultimately in dependence on cathartics.

BIBLIOGRAPHY

ÅGREN G H JAGFELDT AND H BERGLUND The secretin test of pancreatic function in the diagnosis of pancreatic disease *Acta Med Scand* 90 224-271 1936

ALMY T P Alterations in colonic function in man under stress *Gastroenterology* 8 616-626 1947

ALMY T P Alterations in colonic function in man under stress *Gastroenterology* 12 425-436 1949a

ALMY T P Alterations in colonic function in man under stress *Gastroenterology* 12 437-449 1949b

ALMY T P Alterations in colonic function in man under stress *Gastroenterology* 13 93-103 1950

ALVAREZ W C Differences in the behavior of segments from different parts of the intestine *Amer J Physiol* 45 512-550 1918a

ALVAREZ W C AND L STARKWEATHER The metabolic gradient underlying intestinal peristalsis *Amer J Physiol* 46 186-208 1918b

ALVAREZ W C AND K HOSOI A gradient of irritability in the small intestine *Amer J Physiol* 89 182-187 1929

ALVAREZ W C AND L STARKWEATHER The metabolic gradient underlying colonic peristalsis *Amer J Physiol* 47 293-301 1918

AMBACHI N The electrical activity of isolated mammalian intestines *J Physiol (London)* 106 139-153 1947

ANAND B K AND J R BROBECK Localization of a feeding center in the hypothalamus of the rat *Proc Soc Exp Biol Med* 77 323-324 1951a

ANAND B K AND J R BROBECK Hypothalamic control of food intake in rats and cats *Yale J Biol Med* 24 123-140 1951b

ANDERSSON S G NILSON AND B UVNAS Mechanisms of duodenal inhibition of gastric acid secretion In *Seventeenth Nobel Symposium Frontiers of Gastrointestinal Hormone Research* edited by S Andersson Stockholm Almqvist & Wiksells 1970

BABKIN B P *Secretory Mechanism of the Digestive Glands* Ed 2 New York Hoeber 1950 p 664-854

BALL E G The composition of pancreatic juice and blood serum as influenced by the injection of inorganic salts *J Biol Chem* 86 449-462 1930

BALO J AND I BANGA Die Zerstörung der elastischen Fasern der Gefässwand *Schweiz Z Path Bact* 12 350-362 1949

BARCLAY A E *The Digestive Tract A Radiological Study of its Anatomy Physiology and Pathology* London MacMillan 1933 395 pp

BARCROFT J *The Respiratory Function of the Blood* London Cambridge Univ Press 1914 320 pp

BARD P A diencephalic mechanism for the expression of rage with special reference to the sympathetic nervous system *Amer J Physiol* 84 490-513 1928

BARD P The central representation of the sympathetic system as indicated by certain physiologic observations *Arch Neurol Psychiat* 22 230-246 1929

BASS P C F CODE AND E H LAMBERT Motor and electric activity of the duodenum *Amer J Physiol* 201 287-291 1961

BAYLISS W M AND E H STARLING The movements and innervation of the small intestine *J Physiol (London)* 24 99-143 1899

BAYLISS W M AND D H STARLING The movements and innervation of the large intestine *J Physiol (London)* 26 107-118 1901

BAYLISS W M AND E H STARLING The mechanism of pancreatic secretion *J Physiol (London)* 28 325-353 1902

BEAUMONT W *Experiments and Observations on the Gastric Juice and the Physiology of Digestion* Plattsburgh New York F P Allen 1833 (Facsimile edition reprinted on the occasion of the XIIIth International Physiological Congress Boston 1929)

BEHER W T G D BAKER AND W L ANTHONY Feedback control of cholesterol biosynthesis in the mouse *Proc Soc Exp Biol Med* 109 863-868 1962

BENCE JONES H Contributions to the chemistry of the urine Paper III Part I On the variations of the acidity of the urine in the state of health *Phil Trans Roy Soc London* CXL 669-679 1850

BENTLEY F H AND R H SMITHWICK Visceral pain produced by balloon distension of the jejunum *Lancet* 2 389-391 1940

BERGER E H Distribution of parietal cells in the stomach a histotopographic study *Amer J Anat* 54 87-114 1934

BERGSTROM S Bile acids formation and metabolism In *Ciba Foundation Symposium Biosynthesis of Terpenes and Sterols* edited by G E W Wolstenholme and M O'Connor Boston Little Brown 1959 p 185-203

BERGSTROM S R BLOMSTRAND AND B BORGSTROM Route of absorption and distribution of oleic acid and triolein in the rat *Biochim J* 58 600-604 1954

BERNARD C De l'influence de deux ordres de nerfs qui déterminent les variations de couleur du sang veineux dans les organes glanduliformes *C R Acad Sci [D] (Paris)* 41 241-254 1855

BERNARD C Du rôle des actions réflexes paralytiques dans les phénomènes des sécrétions *J Anat* 1 307-314 1864

BIGGS M W D KRITCHEVSKY D COLMAN J W GOFMAN H B JONES E F LINDGREN G HYDE AND J P LYON Observations on the fate of ingested cholesterol in man *Circulation* 6 559-566 1952

BIHLER I K V HAWKINS AND R K CRANE Studies on the mechanism of intestinal absorption of sugars VI The specificity and other properties of Na dependent entrance of sugars into intestinal tissue and under anaerobic conditions in vitro *Biochim Biophys Acta* 59 94-102 1962

BILLING B H P G COLE AND G H LATHE The excretion of bilirubin as a diglucuronide giving the direct van den Bergh reaction *Biochem J* 65 774-784 1957

BLACKBURN C M AND C F CODE Inhibitors of gastric secretion occurring in gastric juice and gastric mucin *Amer J Physiol* 155 197-1948

BLICKENSTAFF D M I GROSSMAN AND A C IVY Stimulating effect of sugar fat and meat meals on duodenal secretion in the dog *Amer J Physiol* 158 122-128 1949

BOEHRER B AND E A DE LA LANDE Electrophoresis of human saliva and identification of inherited variants of amylase isozymes *Aust J Exp Biol Med Sci* 47 97-103 1969

BORGSTRÖM B Studies on intestinal cholesterol absorption in the human *J Clin Invest* 39 809-815 1960

BORGSTROM B A DAHLQVIST G LUNDH AND J SJOVALL Studies of intestinal digestion and absorption in the human *J Clin Invest* 36 1521-1536 1957

BOSMA J F Deglutition pharyngeal stage *Physiol Rev* 37 275-300 1957

BOYDEN E A An analysis of the reaction of the human gall bladder to food *Anat Rec* 40 147-192 1928

BOZLER E The action potentials of visceral smooth muscle *Amer J Physiol* 124 502-510 1938

BOZLER E Electrophysiological studies on the motility of the gastrointestinal tract *Amer J Physiol* 127 301-307 1939

BOZLER E The activity of the pacemaker previous to the discharge of a muscular impulse *Amer J Physiol* 136 543-552 1942

BOZLER E The action potentials of the stomach *Amer J Physiol* 144 693-700 1945

BOZLER E Reflex peristalsis of the intestine *Amer J Physiol* 157 338-342 1949

BRAUER R W Mechanisms of bile secretion *Gastroenterology* 34 1021-1028 1958

BROBECK J R Neural basis of hunger appetite and satiety *Gastroenterology* 32 169-174 1957

BROBECK J R J TEPPERMAN AND C N H LONG Experimental hypothalamic hyperphagia in the albino rat *Yale J Biol Med* 15 831-853 1943

BRO RASMUSSEN F S A KILLMANN AND J H THAYSEN The composition of pancreatic juice as compared to sweat, parotid saliva and tears *Acta Physiol Scand* 37 97-113, 1956

BRUNSCHWIG A J VAN PROHASKA T H CLARKE, AND E KANDEL A secretory depressant in gastric juice of patients with pernicious anemia *J Clin Invest* 18 415-422 1939

BURGEN, A S V AND N G EMMELIN Physiology of the salivary glands In *Monographs of the Physiological Society 8* London Arnold 1961

BURGEN A S V AND P SEEMAN The role of the salivary duct system in the formation of the saliva *Canad J Biochem* 36 119-143, 1958

BURTON P D G EVANS A A HARPER H T HOWAT S OLEESKY J E SCOTT AND H VARLEY Test of pancreatic

function in man based on the analysis of duodenal contents after administration of secretin and pancreozymin *Gut* 1 111-124 1960
CALDWELL F T JR LEVITSKY AND B ROSENBERG Dietary induction and dissolution of cholesterol gallstones in the mouse *J Lab Clin Med* 63 188 437 1964
CANNON W B The movements of the intestines studied by means of roentgen rays *Amer J Physiol* 6 251-277 1902
CANNON W B The passage of different foodstuffs from the stomach and through the small intestine *Amer J Physiol* 12 387-418 1904
CANNON W B The nature of gastric peristalsis *Amer J Physiol* 29 250-266 1911a
CANNON W B *The Mechanical Factors of Digestion* London Arnold 1911b
CANNON W B Peristalsis segmentation and the myenteric reflex *Amer J Physiol* 30 114-128 1912
CANNON W B *Bodily Changes in Pain Hunger Fear and Rage* New York Appleton 1915
CANNON W B Hunger and thirst In *Foundations of Experimental Psychology* edited by C Murchison Worcester Mass Clark Univ Press 1929 p 434-448
CANTAROW A AND B SCHEPARTZ *Biochemistry* Philadelphia Saunders 1954
CAPPS J A AND G H COLEMAN *An Experimental and Clinical Study of Pain in the Pleura Pericardium and Peritoneum* New York Macmillan 1932
CARD W I AND I N MARKS The relationship between the acid output of the stomach following the 'maximal' histamine stimulation and the parietal cell mass *Clin Sci* 19 147-163 1960
CARLSON A J Contributions to the physiology of the stomach I The character of the movements of the empty stomach in man *Amer J Physiol* 31 151-168 1912
CARLSON A J Contributions to the physiology of the stomach II The relation between the contractions of the empty stomach and the sensation of hunger *Amer J Physiol* 31 175-192 1913a
CARLSON A J Contributions to the physiology of the stomach III The contractions of the empty stomach inhibited reflexly from the mouth *Amer J Physiol* 31 212-222 1913b
CARLSON A J Contributions to the physiology of the stomach IV The influence of the contractions of the empty stomach in man on the vasomotor centre on the rate of the heart beat and on the reflex excitability of the spinal cord *Amer J Physiol* 31 318-327 1913c
CARLSON A J T E BOYD AND J F PEARCY Studies on the visceral sensory nervous system *Arch Intern Med (Chicago)* 30 409-433 1922a
CARLSON A J T E BOYD AND J F PEARCY Studies on the visceral sensory nervous system *Amer J Physiol* 61 14-41 1922b
CARLSON A J AND L H BRAAFLADT Contributions to the physiology of the stomach XVIII On the sensibility of the gastric mucosa *Amer J Physiol* 36 153-170 1915
CODE C F C M BLACKBURN G R LIVERMORE AND H V RATKE A method for the quantitative determination of gastric secretory inhibition *Gastroenterology* 13 573-588 1949
CODE C F AND H C CARLSON Motor activity of the stomach In *Handbook of Physiology* edited by C F Code et al Washington Amer Physiol Soc 1968 Sect 6 Vol IV p 1903-1916
CODE C F N C HIGHTOWER JR AND C G MORLOCK Motility of the alimentary canal in man *Amer J Med* 13 328-351 1952
CODE C F G R WILKINSON AND W G SAUER Normal and some abnormal colonic motor patterns in man *Ann N Y Acad Sci* 58 317-335 1954
COLE L G Motor phenomena of the stomach pylorus and cap observed roentgenographically *Amer J Physiol* 42 618-619 1917
CORI C F Mammalian carbohydrate metabolism *Physiol Rev* 11 143-275 1931
COX A J Stomach size and its relation to chronic peptic ulcer *Arch Path (Chicago)* 54 407-422 1952
CRANE R K Hypothesis for mechanism of intestinal active transport of sugars *Fed Proc* 21 891-895 1962
CRANE, R K Na dependent transport in the intestine and other animal tissues *Fed Proc* 24 1000-1006 1965
CRANE R K AND P MANDELSTAM The active transport of sugars by various preparations of hamster intestine *Biochem Biophys Acta* 45 460-476 1960
CRANE R K D MILLER AND I BIHLER The restrictions on possible mechanisms of intestinal active transport of sugars In *Membrane Transport and Metabolism Symposium* edited by A Kleinzeller and A Kotyk Prague Publishing House of the Czechoslovak Academy of Science 1961 p 439-449
CRIGLER J F JR AND V A NAJJAR Congenital familial nonhemolytic jaundice with kernicterus *Pediatrics* 10 169-180 1952
CROSBY W H A concept of the pathogenesis of anemia applied to disorders of the intestinal mucosa *Amer J Dig Dis* 6 492-498 1961
CURRAN P F AND S G SCHULTZ Intestinal absorption of sodium chloride and water In *Handbook of Physiology* edited by C F Code et al Washington Amer Physiol Soc 1968 Sect 6 Vol III p 1245-1275
DAHLQVIST A The intestinal disaccharidases and disaccharide intolerance *Gastroenterology* 43 694-696 1962a
DAHLQVIST A Specificity of human intestinal disaccharidases and implications for hereditary disaccharide intolerance *J Clin Invest* 41 463-470 1962b
DAM H AND F CHRISTENSEN Alimentary production of gall stones in hamsters *Acta Path Microbiol Scand* 30 236-242 1952
DANIELSSON H Present status of research on catabolism and excretion of cholesterol *Advances Lipid Res* 1 335-385 1963
DANIELSSON H K FINARSSON AND G JOHANSSON Effect of biliary drainage on individual reactions in the conversion of cholesterol to taurocholic acid *Europ J Biochem* 2 44-49 1967
DAVENPORT H W Gastric carbonic anhydrase *J Physiol (London)* 97 32-43 1939
DAVENPORT H W Metabolic aspects of gastric acid secretion In *Metabolic Aspects of Transport across Cell Membranes* edited by Q R Murphy Madison Univ Wisconsin Press 1957 p 295-302
DAVENPORT H W *Physiology of the Digestive Tract* Ed 2 Chicago Year Book 1966
DAVIES R E The mechanism of hydrochloric acid production by the stomach *Biol Rev* 26 87-120 1951
DAVIES R E AND V M LONGMUIR Production of ulcers in isolated frog gastric mucosa *Biochem J* 42 621-627 1948
DEBEER E J C G JOHNSON AND D W WILSON The composition of intestinal secretions *J Biol Chem* 108 113-120 1935
DEUTSCH W AND H D RAPER The respiration and metabolism of submaxillary gland tissue of the cat *J Physiol (London)* 92 439-458 1938
DIAMOND J M The reabsorptive function of the gall bladder *J Physiol (London)* 161 442-473 1962a
DIAMOND J M The mechanism of solute transport by the gall bladder *J Physiol (London)* 161 474-502 1962b
DIAMOND J M The mechanism of water transport by the gall bladder *J Physiol (London)* 161 503-527 1962
DREILING D A AND H D JANOWITZ The secretion of electrolytes by the human pancreas the effect of Diamox ACTH and disease *Amer J Dig Dis* 4 137-144 1959
DREILING D A AND H D JANOWITZ Measurement of pancreatic secretory function In *The Exocrine Pancreas* edited by A V S de Reuck and M P Cameron Ciba Foundation Symposium Boston Little Brown 1962 p 225-258
DREILING D A H D JANOWTIZ AND C V PERRIER Pancreatic inflammatory disease a physiologic approach New York Hoeber 1964
DUBIN I N AND F B JOHNSON Chronic idiopathic jaundice with unidentified pigment in liver cells a new clinicopathologic entity with report of 12 cases *Medicine (Balt)* 33 155-197 1954
DWORKEN H J Recent experiences with spontaneously disappearing gallstones *Gastroenterology* 38 76-86 1960
EDKINS J S The chemical mechanism of gastric secretion *J Physiol (London)* 34 033-144 1906
ELIAS H A re examination of the structure of the mammalian liver I Parenchymal architecture *Amer J Anat* 84 311-334 1949a
ELIAS H A re examination of the structure of the mammalian liver II The hepatic lobule and its relation to the vascular and biliary system *Amer J Anat* 85 379 1949b
ELIAS H Morphology of the liver In *Liver Injury* edited by F W Hoffbauer Josiah New York Transactions of the Eleventh Conference Macy Foundation 1953 p 111-199
EMMELIN N Paralytic secretion of saliva An example of super

sensitivity after denervation *Physiol Rev* 32 21 46 1952
ERSPAMER V AND B ASERO Identification of enteramine the specific hormone of the enterochromaffin cell system as 5 hydroxytryptamine *Nature (London)* 169 800 801 1952
EVETT R D J A HIGGINS AND A I BROWN JR The fine structure of normal mucosa in human gallbladder *Gastroenterology* 47 49 60 1964
FARRELL J I AND A C IVY Studies on the motility of the transplanted gastric pouch *Amer J Physiol* 46 227 238 1936
FISHER R B AND J N HUNT The inorganic components of gastric secretion *J Physiol (London)* 111 138 149 1950
FISHER R B AND D S PARSONS Glucose absorption from surviving rat small intestine *J Physiol (London)* 110 81 90 1949
FLOREY H The secretion of mucus by the colon *Brit J Exp Path* 11 318 361 1930
FLOREY H W *and* H F HARDING The function of Brunner gland and the pyloric end of the stomach *J Path Bact* 38 131 45 1933
FLOREY H W AND H L HARDING Further observations on the secretion of Brunner glands *J Path Bact* 39 [illegible] 46 1934
FLOREY H W AND H F HARDING The nature of the hormone controlling Brunner gland *Quart J Exp Physiol* 26 [illegible] 1935a
FLOREY H W AND H F HARDING A humoral control of the secretion of Brunner's gland *Proc Roy Soc London [Biol]* 111 [illegible] 1935b
FLOREY H W R D WRIGHT AND M A JENNINGS The secretion of the intestine *Physiol Rev* 21 [illegible] 1941
FORDTRAN J S AND F F INGELFINGER Absorption of water electrolytes and sugar from the human gut In *Handbook of Physiology* edited by C F Code et al Washington Amer Physiol Soc 1966 sect 6 vol III p 1435 1460
FORDTRAN J S J C RECTOR JR AND N W CARTER The mechanism of sodium absorption in the human small intestine *J Clin Invest* 47 884 900 1968
FRIDHANDLER L AND J H QUASTEL Absorption of amino acids from isolated surviving intestine *Arch Biochem* 36 424 440 1952
FRIEDMAN G B S WOLF J D WAYE AND H D JANOWITZ Correlation of cineradiographic and intraluminal pressure changes in the human duodenum an analysis of the functional significance of nonpropagated waves *Gastroenterology* 49 51 59 1965
FRITZMAN M H F AND W J SNAPE Discoordination of secretion of pancreatic enzyme and bicarbonate in patients with chronic pancreatitis *Gastroenterology* 17 886 897 1950
FRITZ M F AND F P BROOKS Control of bile flow in the chole cystectomized dog *Amer J Physiol* 204 843 849 1963
FULTON J F AND F D INGRAHAM Emotional disturbances following experimental lesions of the base of the brain (prosencephalon) *Amer J Physiol* 90 343 1929
FYKE F E JR AND C F CODE Resting and deglutition pressures in the pharyngo esophageal region *Gastroenterology* 39 24 34 1960
FYKE F E JR, C F CODE AND J F SCHLEGEL The pattern of esophageal sphincter in healthy human beings *Gastroenterology* 30 135 150 1956
GARTNER L M AND I M ARIAS Formation transport metabolism and excretion of bilirubin *New Eng J Med* 280 1439 1445, 1969
GILMAN, A The relation between blood osmotic pressure fluid distribution and voluntary water intake *Amer J Physiol* 120 323 328 1937
GRACE, W J S G WOLF AND H G WOLFF Influence of emotions and feeling states on the behaviour of the human colon *Amer J Physiol* 155 439 440 1948
GRACE W J, S WOLF, AND H G WOLFF *The Human Colon An Experimental Study Based on Direct Observation of Four Fistulous Subjects* New York Hoeber 1951
GREER, M A Suggestive evidence of a primary "drinking center" in the hypothalamus of the rat *Proc Soc Exp Biol Med* 89 59 62 1955
GREGORY, R A., AND H J. TRACY The constitution and properties of two gastrins extracted from hog antral mucosa *Gut* 5 103 114, 1964a.
GREGORY, R A., AND H J TRACY A note on the nature of the gastrin-like stimulant present in Zollinger Ellison tumors *Gut* 5 115 117, 1964b
GROSSBERG A L S A KOMAROV AND H SHAY Distribution of proteins and enzymatic activities in electrophoretic components of canine pancreatic juice *Amer J Physiol* 168 269 282, 1952
GROSSMAN M I Integration of current views on the regulation of hunger and appetite *Ann N Y Acad Sci* 63 76 89 1955
GROSSMAN M I Integration of neural and hormonal control of gastric secretion *Physiologist* 6 [illegible] 1963
GROSSMAN M I The spectrum of gastrointestinal hormone action In *Sixteenth Nobel Symposium Frontiers of Gastrointestinal Hormone Research* edited by S Anderson Stockholm Almqvist & Wiksell 1970
GROSSMAN M I AND A C IVY Effect of ileum upon external secretion of the pancreas *Proc Soc Exp Biol Med* 63 62 63 1946
GROSSMAN M I H D JANOWITZ H RAWSON AND K S KIM The effect of secretin on bile formation in man *Gastroenterology* 13 [illegible] 1949
GROSSMAN S P Eating and drinking elicited by direct adrenergic or cholinergic stimulation of hypothalamus *Science* 132 301 302 1960
HALL E M J GUERRY AND N C HIGHTOWER Comparison of the electrophoretic characteristics of human pancreatic isoamylases with human salivary amylase *Clin Res* 17 103 1969
HAMMARSTEN E F JORPES AND G ÅGREN Versuche zur Resorption von Sekretin *Biochem Z* 264 [illegible] 1933
HAMMARSTEN E O WHANDER AND G ÅGREN Versuche zur Resorption von Sekretin *Acta Med Scand* 68 239 247 1928
HARPER A A AND H S RAPER Pancreozymin a stimulant of the secretion of pancreatic enzymes in extract of the small intestine *J Physiol (London)* 143 145 1943
HARVEY W A second disquisition to John Riolan In *On the Motion of the Heart and Blood in Animals* (Willis translation with two letters addressed to John Riolan) Chicago Regnery 1962 p 166–91
HEPTINSTALL R W AND S W RANSON Hypothalamic lesion and adipose tissue in the rat *Anat Rec* 158 149 152 1940
HIGHTOWER N C JR The physiology of symptoms 10 Swallowing and esophageal motility *Amer J Dig Dis* 3 569 583 1958
HIGHTOWER N C JR Motility of the alimentary canal of man In *Disturbances in the Gastrointestinal Motility* edited by J A Rider and H C Moeller Springfield Ill Thomas 1959 p 3 61
HIGHTOWER N C JR AND C F CODE The quantitative analysis of antral gastric motility recorded in normal human beings, with a study of the effect of neostigmine *Mayo Clin Proc* 25 697 704 1950
HIGHTOWER N C JR, A M OLSEN AND H F MOERSCH A comparison of the effect of neostigmine methyl sulfate methyl choline chloride (Mecholyl) on esophageal intraluminal pressure in normal persons and patients with cardiospasm *Gastroenterology* 26 597 600 1954
HIGHTOWER N C JR AND M F SALEM The rate of pharyngeal and esophageal transport *J Lab Clin Med* 33 820 1948
HIRSCH G C *Form und Stoffwechsel der Golgi Körper Protoplasma Monographie* 18 Berlin Borntraeger 1939
HIRSCH, G C, P L C U HUNQUEIRA H A ROTHSCHILD AND S R DOHI Die Pankreassaft Sekretion bei der Ratte I Die konstitutionelle irreguläre Hungersekretion und ihre Ursachen *Pflügers Arch* 264 [illegible] 1957
HOELZEL F The effect of variations in protein intake on the acidity of the secretion of the fasting stomach *Amer J Physiol* 77 166 180 1926
HOFMANN A F A physicochemical approach to the intraluminal phase of fat absorption *Gastroenterology* 41 56 64 1966
HOFMANN A F Functions of bile in the alimentary canal edited by C F Code et al Washington Amer Physiol Soc 1968a, sect 6 vol V p 2567 2585
HOFMANN, A F Intraluminal factors in the absorption of glycerides In *Medium Chain Triglycerides* edited by J R Senior Philadelphia Univ Pennsylvania Press 1968b p 9 19
HOFMANN A F AND B BORGSTRÖM The intraluminal phase of fat digestion in man the lipid content of the micellar and oil phases of intestinal content obtained during fat digestion and absorption *J Clin Invest* 43 247 257 1964
HOGBEN, C A M The alimentary tract *Ann Rev Physiol* 22 381 406, 1960)

HOKIN L F The synthesis and secretion of amylase by pigeon pancreas in vitro *Biochem J* 48 320–326 1951

HOKIN L E Amino acid requirements of amylase synthesis by pigeon pancreas slices *Biochem J* 50 216–220 1952

HOKIN L E AND M R HOKIN The actions of pancreozymin in pancreas slices and the role of phospholipids in enzyme secretion *J Physiol (London)* 132 442–453 1956

HOKIN L E AND M R HOKIN The mechanism of phosphate exchange in phosphatidic acid in response to acetylcholine *J Biol Chem* 234 1387–1390 1959

HOLLANDER F Studies in gastric secretion IV Variations in the chlorine content of gastric juice and their significance *J Biol Chem* 97 585–604 1932

HOLLANDER F The insulin test for the presence of intact nerve fibers after vagal operations for peptic ulcer *Gastroenterology* 7 607–614 1946

HOLLANDER F AND G R COWGILL Studies in gastric secretion I Gastric juice of constant acidity *J Biol Chem* 91 151–182 1931

HOLMES J H AND M I GREGERSON Relation of salivary flow to the thirst produced in man by intravenous injection of hypertonic salt solution *Amer J Physiol* 151 252–257 1947

HOUCK J C AND Y M PATEL The collagenolytic activity of the pancreas *Ann N Y Acad Sci* 93 333–349 1962

HOWELL A H AND R S MANLY An electronic strain gauge for measuring oral forces *J Dent Res* 27 705–712 1948

HUNT J N Some properties of an alimentary osmoreceptor mechanism J Physiol (*London*) 132 267–288 1956

HUNT J N AND W R SPURRELL The pattern of emptying of the human stomach *J Physiol (London)* 113 157–168 1951

HURST A F *The Sensibility of the Alimentary Canal* London Frowde 1911

IHRE B Human gastric secretion A quantitative study of gastric secretion in normal and pathological conditions *Acta Med Scand* 95 (Suppl XCV) 1938

INGELFINGER F J P KRAMER AND G C SANCHEZ The gastroesophageal vestibule its normal function and its role in cardiospasm and gastroesophageal reflux *Amer J Med Sci* 228 417–425 1954

ISSELBACHER K J AND J R SENIOR The intestinal absorption of carbohydrate and fat *Gastroenterology* 46 287–298 1964

IVY A C M I GROSSMAN AND W H BACHRACH *Peptic Ulcer* New York Blakiston 1950

IVY A C AND E OLDBERG A hormone mechanism for gall bladder contraction and evacuation *Amer J Physiol* 86 599–613 1928

JAMES A H The physiology of gastric digestion In *Monographs of the Physiological Society 4* London Arnold 1957

JANOWITZ H D Pancreatic secretion of fluid and electrolytes In *Handbook of Physiology* edited by C F Code et al Washington Amer Physiol Soc 1967 sect 6 vol II p 925–933

JANOWITZ H D AND D A DREILING The pancreatic secretion of fluid and electrolytes Pages 115–137 In *The Exocrine Pancreas* edited by A V S de Reuck and M P Cameron Ciba Foundation Symposium Boston Little Brown 1962 p 115–137

JANOWITZ H D D A DREILING H L ROLBIN, AND F HOLLANDER Inhibition of the formation of hydrochloric acid in the human stomach by Diamox *Gastroenterology* 33 378–384 1957

JANOWITZ H D AND M I GROSSMAN Some factors affecting the food intake of normal dogs and dogs with esophagostomy and gastric fistula *Amer J Physiol* 159 143–148 1949

JANOWITZ H D AND F HOLLANDER The time factor in the adjustment of food intake to varied caloric requirement in the dog a study of the precision of appetite regulation *Ann N Y Acad Sci* 63 56–67 1955

JANOWITZ H D F HOLLANDER D ORRINGER M H LEVY A WINKELSTEIN M R KAUFMAN AND S G MARGOLIN A quantitative study of the gastric secretory response to sham feeding in a human subject *Gastroenterology* 16 104–116 1950

JOHNSON L R AND M I GROSSMAN Analysis of inhibition of acid secretion by cholecystokinin in dogs *Amer J Physiol* 218 550–554 1970

JOHNSTON J M Mechanism of fat absorption In *Handbook of Physiology* edited by C F Code et al Washington Amer Physiol Soc 1968 sect 6 vol III p 1353–1375

JONES R S , AND F P BROOKS The pyloric antrum as a mediator of insulin induced cholerisis *Physiologist* 8 202 1965

JONES R S K C POWELL AND F P BROOKS The role of the gastric antrum in the control of bile flow *Surg Forum* 16 386–389 1965

JUNIPER K Chemical and ultracentrifugal studies on the macromolecular complex in normal bile of the human gall bladder *Amer J Surg* 107 371–379 1964

KEITH A The nature and anatomy of enteroptosis (Glenard's disease) *Lancet* 1 631 709 1903

KELLEY M L JR E A GORDON AND J A DEWEESE Pressure studies on the ileocolonic junctional zone of dogs *Amer J Physiol* 209 333–339 1965

KINSELLA V J *The Mechanism of Abdominal Pain* Sydney Australia Australasian Medical Publishing Co 1948

KINTER W B Autoradiographic study of intestinal transport In *Proceedings of the Twelfth Annual Conference on Nephrotic Syndrome* edited by J Metcoff New York National Kidney Disease Foundation 1961 p 59

KIYASU J Y B BLOOM AND I L CHAIKOFF The portal transport of absorbed fatty acids *J Biol Chem* 199 415–419 1952

KOCOUR E J AND A C IVY The effect of certain foods on bile volume output recorded in the dog by a quantitative method *Amer J Physiol* 122 325–346 1937

KOSAKA T AND R K S LIM Demonstration of a humoral agent in fat inhibition of gastric secretion *Proc Soc Exp Biol Med* 27 890–891 1930

KOSAKA T AND R K S LIM On the mechanism of the inhibition of gastric motility by fat An inhibitory agent from the intestinal mucosa *Chin J Physiol* 7 5–11 1933

KRAMER P AND F J INGELFINGER Esophageal sensitivity to Mecholyl in cardiospasm *Gastroenterology* 19 241–251 1951

KUNITZ M AND J H NORTHROP Isolation crystallization and general properties of a new proteolytic enzyme and its precursor *J Gen Physiol* 18 433–458 1934

KUNTZ A *The Autonomic Nervous System* Ed 4 Philadelphia Lea & Febiger 1953

LAGERLOF H The secretin test of pancreatic function *Quart J Med* 8 115–126 1939

LARSSON S On the hypothalamic organization of the nervous mechanism regulating food intake *Acta Physiol Scand* 32 (Suppl 115) 1954

LEBLOND C P AND C E STEVENS The constant renewal of the intestinal epithelium in the albino rat *Anat Rec* 100 357–377 1948

LEBLOND C P AND B E WALKER Renewal of cell populations *Physiol Rev* 36 255–276 1956

LERCHE W *The Esophagus and Pharynx in Action A Study of Structure in Relation to Function* Springfield Ill Thomas 1950

LESTER R The Intestinal phase of bile pigment excretion *Gastroenterology* 47 424–427 1964

LESTER R AND R SCHMID Intestinal absorption of bile pigments I The enterohepatic circulation of bilirubin in the rat *J Clin Invest* 42 736–746 1963a

LESTER R AND R SCHMID Intestinal absorption of bile pigments II Bilirubin absorption in man *New Eng J Med* 269 178–182 1963b

LESTER R AND R SCHMID Enterohepatic circulation of urobilinogen *Nature (London)* 201 711–712 1964

LICHTENSTEIN M E AND A C IVY The function of the "valves" of Heister *Surgery* 1 38–52 1937

LIM R K S Observations on the mechanism of the inhibition of gastric function by fat *Quart J Exp Physiol* 23 263–268 1933

LINDSTEDT S The turnover of cholic acid in man bile acids and steroids 51 *Acta Physiol Scand* 40 1–9 1957

LUCKHARDT, A B H T PHILLIPS AND A J CARLSON Contributions to the physiology of the stomach II The control of the pylorus *Amer J Physiol* 50 57–66 1919

LUDWIG C Neue Versuche über die Beihilfe der Nerven zur Speichelabsonderung *Z Rationelle Med* N F 1 255–277 1851

LUNDBERG A Electrophysiology of the submaxillary gland of the cat *Acta Physiol Scand* 35 1–25 1955

LUNDBERG A Secretory potentials in the sublingual gland of the cat *Acta Physiol Scand* 40 21–34 1957

LUNDBERG A Electrophysiology of salivary glands *Physiol Rev* 38 21–40 1958

MACKENZIE J *Symptoms and Their Interpretations* London Shaw 1918

MAGENDIE F *Precis Elementaire de Physiologie* Ed 5 translated by John Revere (An elementary treatise on human physiology) New York Harper 1847

MAKHLOUF G M AND A I BLUM An assessment of pancreatic electrolyte secretion in man and dog *Gastroenterology* 58 972 1970

MANN F C AND T B BOLLMAN The relation of gall bladder to the development of jaundice following obstruction of common bile duct *J Lab Clin Med* 10 540 543 1925

MANN F C T B BOLLMAN AND I R MAGATH Studies on the physiology of the liver IX The formation of bile pigment after total removal of the liver *Amer J Physiol* 69 393 409 1924

MANN F C C SHEARD T B BOLLMAN AND E J BALDES The rate of the formation of bilirubin *Amer J Physiol* 74 197 210 1925

MANN F C C SHEARD T B BOLLMAN AND E J BALDES The formation of bile pigment from hemoglobin *Amer J Physiol* 76 506 514 1926

MARKS I N S A KOMAROV AND H SHAY Maximal acid secretory response to histamine and its relation to parietal cell mass in the dog *Amer J Physiol* 199 79 588 1960

MARKWALD M Uber die Ausbreitung der Erregung und Hemmung von Schluckcentrum auf das Athemcentrum *Z Biol N F* 1 1 1889

MAYER J Regulation of energy intake and the body weight the glucostatic theory and the lipostatic hypothesis *Ann N Y Acad Sci* 63 15 43 1955a

MAYER J The physiological basis of obesity and leanness *Nutr Abstr Rev* 25 597 611 1955b

McCLURE F J Domestic water and dental caries III Fluorine in human saliva *Amer J Dis Child* 62 42 1941

McNALLY E F J F KELLY AND F INGELFINGER Mechanism of belching effects of gastric distention with air *Gastroenterology* 46 234 240 1964

MENGUY R B G A HALTENBECK J T BOLLMAN AND H H GRINDLEY Intraductal pressures and sphincteric resistance in common pancreatic and biliary duct after various stimuli *Surg Gynec Obstet* 106 306 310 1958

MILLER D AND R K CRANE The digestive function of the epithelium of the small intestine I An intracellular locus of disaccharide and sugar phosphate ester hydrolysis *Biochim biophys Acta* 47 281 293 1961

MILLER M C Spontaneous discoloration of gall stone *Gastroenterology* 31 588 591 1956

MILLER T G AND W O ABBOTT Intestinal intubation a practical technique *Amer J Med Sci* 187 89 99 1934

MILTON G W AND A W M SMITH The pacemaking area of the duodenum *J Physiol (London)* 127 100 114 1956

MIYAKAWA K J KIRSNER AND A R MIA Pressures in mouth and pharynx during sucking and swallowing *Proc Proc* 16 89 1957

MONTGOMERY M L G E SHILLING AND T T CHAIKOFF Elimination of sodium in pancreatic juice as measured by radioactive sodium *Amer J Physiol* 141 548 551 1944

MORGANE P J Distinct feeding and hunger motivating systems in the lateral hypothalamus of the rat *Science* 133 887 888 1961

MORRISON S D R J BARRNETT AND J MAYER Localization of lesions in the lateral hypothalamus of rats with induced adipsia and aphagia *Amer J Physiol* 194 230 234 1958

MORRISON S D AND J MAYER Adipsia and aphagia in rats after lateral subthalamic lesions *Amer J Physiol* 191 248 254 1957

MYANT N B AND H A EDER The effect of biliary drainage upon the synthesis of cholesterol in the liver *J Lipid Res* 2 363 368 1961

NASSET F S Enterocrinin a hormone which excites the glands of the small intestine *Amer J Physiol* 121 481 487 1938

PRESHAW R M Integration of nervous and hormonal mechanisms for external pancreatic secretion In *Handbook of Physiology* edited by C F Code et al Washington Amer Physiol Soc 1967 sect 6 vol II p 997-1076

NASSET E S AND A A PARRY Passage of fluid and certain dissolved substances through the intestinal mucosa as influenced by changes in hydrostatic pressure *Amer J Physiol* 109 615 625 1934

NASSET E S P SCHWARTZ AND H V WEISS The digestion of proteins in vivo *J Nutr* 36 85 94 1948

NIAMI K D AND G WISEMAN The transamination of glutamic and aspartic acids during absorption by the small intestine of the dog in vivo *J Physiol (London)* 139 417 430 1957

NEURATH H Proteolytic digestive enzymes *Sci Amer* 211 68-79 1964

NORTHUP D The secretory metabolism of the salivary glands *Amer J Physiol* 114 46 52 1935

OKEY R Gallstone formation and intake of fat vitamins in cholesterol fed guinea pigs *Proc Soc Exp Biol Med* 51 349 350 1942

OLSON E J AND S J GRAY Serotonin and gastroenterology *Amer J Gastroent* 39 280 289 1963

PAINTAL A S A study of gastric stretch receptors Their role in the peripheral mechanism of satiation of hunger and thirst *J Physiol (London)* 126 255 270 1954

PALADE G E P SIEKEVITZ AND L G CARO Structure chemistry and function of the pancreatic exocrine cell In *The Exocrine Pancreas* edited by A V S de Reuck and Margaret P Cameron Ciba Foundation Symposium Boston Little Brown 1962 p 23 49

PAVLOV I P *The Work of the Digestive Glands* Ed 2 Philadelphia Lippincott 1910

POGRUND R S AND E K STEGGERDA Influence of gaseous transfer between the colon and blood stream on percentage gas composition of intestinal flatus in man *Amer J Physiol* 155 475 482 1948

POMMERENKE W T A study of censory area eliciting the swallowing reflex *Amer J Physiol* 84 36 41 1928

PORTER K R Electron microscopy of basophilic components of cytoplasm *J Histochem Cytochem* 2 346 354 1954

POULTON E P Oliver Sharpey lecture on an experimental study of certain gastric secretions *Lancet* 2 1177 1240 1928

QUIGLEY J P A modern explanation of the gastric emptying mechanism *Amer J Dig Dis* 10 418 423 1943

QUIGLEY J P The role of the duodenum in regulating the ingestion of food *Ann N Y Acad Sci* 63 611 1955

QUIGLEY J P AND D H BRODY *Medical Physics* Edited by O Glaser Chicago Year Book 1950 II p 980 993

QUIGLEY J P AND W R HALLARAN The independence of spontaneous gastrointestinal motility and blood sugar levels *Amer J Physiol* 140 103 110 1943

QUIGLEY J P V JOHNSON AND E F SOLOMON Action of insulin on the motility of the gastrointestinal tract I Action on the stomach of normal and decerebrate man *Amer J Physiol* 90 89 98 1929

QUIGLEY J P AND E F SOLOMON Action of insulin on the motility of the gastrointestinal tract V a Action on the human duodenum b Action on the colon of dog *Amer J Physiol* 91 488 496 1930

QUIGLEY J P AND R D TEMPLETON Action of insulin on the motility of the gastrointestinal tract IV Action on the stomach following double vagotomy *Amer J Physiol* 91 484 487 1930

QUIGLEY J P H I ZITTLEMAN AND A C IVY Analysis of factors involved in gastric motor inhibition by fats *Amer J Physiol* 108 643 651 1934

RAVDIN I S C G JOHNSTON J H MUSHIN AND C RIEGEL Studies of gall bladder function IV The absorption of chloride from the bile free gall bladder *Amer J Physiol* 89 648 661 1932a

RAVDIN I S C G JOHNSTON C RIEGEL AND S L WRIGHT JR Studies of gall bladder function VII The anion content of hepatic and gall bladder bile *Amer J Physiol* 100 413 427 1932b

RAWINSON H E The changes in the dividular and demilune cells of the simple and stimulated parotid submaxillary gland of the cat *J Anat* 70 143 148 1935

RAWSON A J Human bile proteins I Proteins identified by antibody to human serum *Clin Chem* 8 310 317 1962

REHM W S A theory of the formation of hydrochloric acid by the stomach *Gastroenterology* 14 401 417 1950

REINHOLD J G AND D W WILSON The acid base composition of hepatic bile *Amer J Physiol* 107 378 387 1934

REISER R M J BRYSON M J CARR AND K A KUIKEN The intestinal absorption of triglycerides *J Biol Chem* 194 131-138 1952

ROHRER G V J R SCOTT W JOFI AND S WOLF The fine

structure of human gastric parietal cells *Amer J Dig Dis* 10 13–21 1965

ROTOR A B L MANAHAN AND A FLORENTIN Familial non hemolytic jaundice with direct van den Bergh reaction *Acta Med Phillipina* 5 37–49 1948

ROUS P AND P D McMASTER The concentrating activity of the gallbladder *J Exp Med* 34 47–73 1921

RUSHMER R F AND J A HENDRON The act of deglutition A cinefluorographic study *J Appl Physiol* 3 622–630 1951

RUSSELL I S AND W BURNETT The proteins of human bile *Gastroenterology* 45 730–739 1963

SAMI OFF I M AND P L TOWNES Heterogeneity and genetic polymorphism of human pepsinogen *Gastroenterology* 56 1194 1969

SANCHEZ G C P KRAMER AND F J INGELFINGER Motor mechanisms of the esophagus particularly of its distal portion *Gastroenterology* 25 321–332 1953

SAUNDERS D R AND A M DAWSON Studies on the metabolism of glycerol by the small intestine in vitro and in vivo *Biochem J* 82 477–483 1962

SCHAFFNER F Morphologic studies on bile secretion *Amer J Dig Dis* 10 99–115 1965

SCHIFFRIN M J AND E S NASSET The response of jejunum and ileum to food and enterocrinin *Amer J Physiol* 128 70–80 1939

SCHMID R Direct reacting bilirubin bilirubin glucuronide in serum bile and urine *Science* 124 76–77 1956

SCHNEYER L H AND L K LEVIN Rate of secretion by individual salivary gland pairs of man under conditions of reduced exogenous stimulation *J Appl Physiol* 7 508–512 1955

SCHOLER J F AND C F CODE Rate of absorption of water from stomach and small bowel of human beings *Gastroenterology* 27 565–577 1954

SCOTT B L AND D C PEASE Electron microscopy of the major salivary glands of the rat *Anat Rec* 127 364 1957

SCOTT W W C C SCOTT AND A B LUCKHARDT Observations on the blood sugar level before during and after hunger periods in humans *Amer J Physiol* 123 243–247 1938

SEDAR A W The fine structure of the oxyntic cell in relation to functional activity of the stomach *Ann N Y Acad Sci* 99 9–29 1962

SEDAR A W AND M H F FRIEDMAN Correlation of the fine structure of the gastric parietal cell (dog) with functional activity of the stomach *J Biophys Biochem Cytol* 11 349–363 1961

SENIOR J R AND K J ISSELBACHER Direct esterification of monoglycerides with palmityl coenzyme A by intestinal epithelial subcellular fractions *J Biol Chem* 237 1454–1459 1962

SHARE I E MARTYNIUK AND M I GROSSMAN Effect of prolonged intragastric feeding on oral food intake in dogs *Amer J Physiol* 169 229–235 1952

SIEGEL C I AND T R HENDRIX The clinical value of esophageal motor studies *Postgrad Med* 29 505–512, 1961

SLUITER J W Das Restitutionsproblem in der Pancreaselle I Die Bedeutung des Golgi Apparates *Z Zellforsch* 33 187 1944

SNELL R S AND J R GARRETT Histochemical appearances of cholinesterase in the submaxillary and sublingual salivary glands of the rat *Nature (London)* 178 1177–1178 1956

SPANNER R Der Abkürzungskreislauf der Glandula submaxillaris *Z Anat Entwicklungsgesch* 107 124–153 1937

STARLING E H cited by Babkin, 1950 *Mercers' Company Lectures on Recent Advances in the Physiology of Digestion* London Constable 1906

STARR H E Studies of human mixed saliva I The determination of the hydrogen ion concentration of human mixed saliva *J Biol Chem* 54 43–64 1922

STEGGERDA F R Observations on the water intake in an adult man with dysfunctioning salivary glands *Amer J Physiol* 132 517–521 1941

STROM G AND B UVNAS Motor responses of gastro intestinal tract and bladder to topical stimulation of the frontal lobe basal ganglia and hypothalamus in the cat *Acta Physiol Scand* 21 90–104 1950

STROMBLAD B C R Oxygen consumption of the normal and denervated submaxillary gland in vitro *Acta Physiol Scand* 40 130–145 1957

STORMONT, D L *The salivary glands* In *Special Cytology* Ed 2, edited by E V Cowdry New York Hoeber 1932 vol 1 p 151–196

STUNKARD A J AND H G WOLFF Studies on the physiology of hunger I The effect of intravenous administration of glucose on gastric hunger contractions in man *J Clin Invest* 35 954–963 1956

SUN D C H AND H SHAY Pancreozymin secretin test The combined study of serum enzymes and duodenal contents in the diagnosis of pancreatic disease *Gastroenterology* 38 570–581 1960

TANTURI C A AND A C IVY On the existence of secretory nerves in the vagi for and the reflex excitation and inhibition of bile secretion *Amer J Physiol* 121 270–283 1938

TEPPERMAN J F T CALDWELL AND H M TEPPERMAN Induction of gallstones in mice by feeding a cholesterol cholic acid containing diet *Amer J Physiol* 206 628–634 1964

THAYSEN J H N A THORN AND I L SCHWARTZ Excretion of sodium potassium chloride and carbon dioxide in human parotid saliva *Amer J Physiol* 178 155–159 1954

THIRY L Eine neue Methode den Dunndarm zu isoleiren *Sitzungsber Kaiser Akad Wiss Math Naturw Classe* 50(1) 77–96 1864

THOMAS J E Digestion and the nervous system A review of the literature *Amer J Dig Dis* 10 201–206 1943

THOMAS J F The gradient theory versus the reflex theory of intestinal peristalsis *Amer J Gastroent* 23 13–25 1955

THOMAS J E AND M H F FRIEDMAN Physiology of upper gastrointestinal tract as it relates to peptic ulcer Chapter 3 In *Peptic Ulcer Clinical Aspects Diagnosis Management* Edited by D J Sandweiss Philadelphia Saunders 1951 p 33–60

THOMAS J E AND H WHEELON The nervous control of the pyloric sphincter *J Lab Clin Med* 7 375–391 1922

TOWBIN E J Thirst and hunger behavior in normal dogs and the effects of vagotomy and sympathectomy *Amer J Physiol* 182 377–382 1955

TREADWELL C R AND G V VAHOUNY Cholesterol Absorption In *Handbook of Physiology* edited by C F Code et al Washington Amer Physiol Soc 1968 sect 6 vol III p 1402–1438

TRIER J S AND C E RUBIN Electron microscopy of the small intestine a review *Gastroenterology* 49 574–603 1965

TRISKA W Experimentelle Studien uber die Beisskraft *Pflueger Arch* 204 660–667 1924

TUTTLE S G A BETTARELLO AND M I GROSSMAN Gastroesophageal regurgitation *J A M A* 176 498–500 1961

VAHOUNY G V C H WOO AND C R TREADWELL Quantitative effects of bile salt and fatty acid on cholesterol absorption in the rat *Amer J Physiol* 193 41–46 1958

VERZAR F AND E McDOUGALL *Absorption from the Intestine* New York Longmans Green 1936

VINEBERG A M The activation of different elements of the gastric secretion by variation of vagal stimulation *Amer J Physiol* 96 363–371 1931

WEIR J F E E LARSON AND L G ROWNTREE Studies in diabetes insipidus water balance and water intoxication *Arch Intern Med (Chicago)* 29 306–330 1922

WHEELER H O Water and electrolytes in bile In *Handbook of Physiology* edited by C F Code et al Washington Amer Physiol Soc 1968 sect 6 vol V p 2409–2431

WHEELER H O AND O L RAMOS Determinants of the flow and composition of bile in the unanesthetized dog during constant infusions of sodium taurocholate *J Clin Invest* 39 161–170 1960

WHIPPLE G H AND C W HOOPER Icterus A rapid change of hemoglobin to bile pigment in the circulation outside the liver *J Exp Med* 17 612–635 1913

WHITE T T G LUNDH AND D F MAGEE Evidence for the existence of a gastropancreatic reflex *Amer J Physiol* 198 725–728 1960

WILSON T H AND G WISEMAN The use of sacs of everted small intestine for the study of the transference of substances from the mucosal to the serosal surface *J Physiol (London)* 123 116–125 1954

WISEMAN G Active sterochemically selective absorption of amino acids from rat small intestine *J Physiol (London)* 114 7P–8P 1951

WISEMAN G Absorption of amino acids using an in vitro technique *J Physiol (London)* 120 63–72 1953

WISEMAN G Absorption from the intestine In *Handbook of Physiology* edited by C F Code et al Washington Amer Physiol Soc 1966 sect 6, vol III p 1277–1307

WOLF S AND H G WOLFF *Human Gastric Function* London Oxford Univ Press 1943

WRIGHT R D H W FLOREY AND M A JENNINGS The secretion of the colon of the cat *Quart J Exp Physiol* 28 207 229 1938

WRIGHT R D M A JENNINGS H W FLOREY AND R H M The influence of nerves and drugs on secretion by the small intestine and an investigation of the enzymes in intestinal juice *Quart J Exp Physiol* 30 73 120 1940

ZAFIRAS AND M I GROSSMAN The effect of gastrin and histamine on secretion of bile *Gastroenterology* 50 300 305 1966

ZETTERQVIST H *The Ultrastructural Organization of the Columnar Absorbing Cells of the Mouse Jejunum* Stockholm Karolinska Institute Aktiebolaget Godvil 1956

ZIMMERMAN K W cited in Babkin 1950 Die Speicheldrusen der Mundhohle und die Bauchspeicheldruse In *Von Mollendorffs Handbuch der Mikroskopischen Anatomie des Menschen* Berlin Springer 1927 vol v part I p 61 214

3

Circulation

3 / THE CARDIOVASCULAR SYSTEM

The blood transport system functions to distribute and to collect blood from diffusion sites within the body It operates within definite limits of pressure and flow required for adequate perfusion of organs and tissues Complex control mechanisms govern the activities of its various parts and their functional interrelations with body tissues

The adult mammalian heart consists of two fluid pumps in series, the right heart supplying blood to the lung circulation, and the left heart to the remainder of the body (systemic circulation) Exacting equalization of the output of these two pumps is required to maintain the disparate blood volume distribution between these two circulations which is compatible with life This coordination must be accomplished despite increases in total cardiac output which can amount to 500% or more as cardiac output adjusts to body needs (e g , rest vs exercise)

The vascular system is an enormous network of viscoelastic tubes Through serial bifurcations the single outflow vessel from the left ventricle divides in the mesenteric vascular bed alone into more than two billion channels with a total cross section over one thousand times greater than the parent vessel, these then rejoin in repeated confluences forming the two venae cavae which return the blood to the right heart

The vascular system can be classified into three general types of vessels with different functions the arteries, a distributing system, the microcirculation, a diffusion and filtration system, and the veins, a collecting system

The grossly visible distributing arteries carry blood to the microcirculation (vessels less than 100 μ in diameter), which includes the capillary beds (actual diffusing portion) From here blood flows into the collecting venous system (see Figs 3 1 and 3 2)

The blood moves through each segment of the branching and reuniting vascular network in equal volume flow along a progressively decreasing pressure gradient (Fig 3 2*A*, *B*, and *C*) in both the pulmonic and systemic circuits The beginning and end of each circuit is the heart, which, through mechanical work and one-way valvular action, provides the energy required to maintain the pressure and flow gradients

The level of arterial blood pressure maintained must be sufficient to insure adequate perfusion of the microcirculation bed against the forces of gravity (upright posture) acceleration, and deceleration (inertial effects), mechanical distortion and pressures on distributing arteries and the tendency for some vessels to close altogether owing to their own wall tension The biological significance of differing blood pressure ranges characteristic of various species has not been systematically analyzed However, certain correlations can be cited For example the resting systemic mean arterial pressures at heart level decrease from giraffe to ox to man (219, 135, 90 mm Hg, respectively) in proportion to the characteristic distance of the brain above the heart (160 64, 34 cm respectively) in the erect posture (Patterson et al , 1965b) The mean arterial pressures at brain level in these three species (obtained by subtracting a correction for pressure exerted by a column of blood equal to the heart-brain distance), would be less disparate (100, 88, 65 mm Hg, respectively) (see Fig 3 3 *A* and *B*) Birds, subject to marked alterations in gravitational, accelerational, and decelerational effects associated with flying, have, as a class, higher arterial pressures than mammals Finally, in mammals and birds the pulmonic circulation, protected from mechanical effects in the rigid thoracic cage and situated at all times close to heart level, operates at a mean arterial pressure about one-seventh that of the large systemic arteries

There is also an interesting relation between plasma colloid osmotic pressure and blood pressures characteristic of various species (Figs 3 4 and Table 3 1)

The microcirculation consists of a system of small arterioles which lead blood into the capillaries, where diffusion and filtration take place, and small venules which carry the blood into

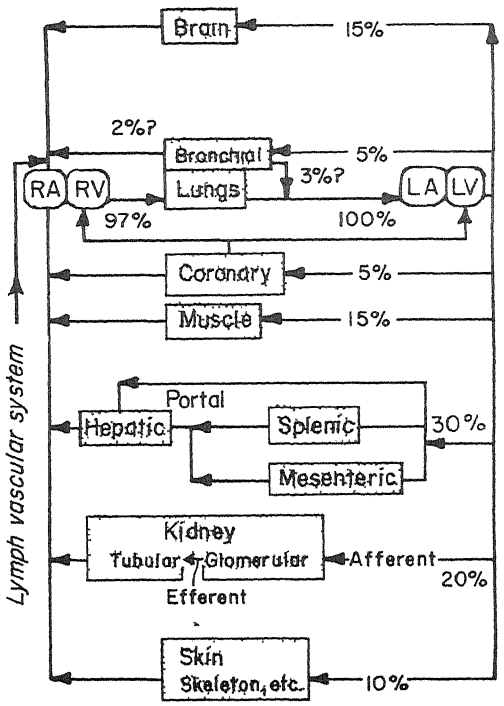


Fig 3 1 Stylized block diagrams of the circulatory system A schema of the circulatory system including the major vascular components (heart arteries microcirculation veins) which distribute to and collect blood from the tissues The special circulatory beds to be discussed separately are indicated and the relationship to the lymphatic collecting duct system represented Solid lines depict arteries and veins arrowheads show the direction of flow and shaded blocks indicate the microcirculation (arterioles capillaries and venules) Approximate percentages of cardiac output going to the various circulatory beds are indicated near each line

the collecting veins At this level, lymph capillaries occur These function as fluid- and particle-collecting vessels They unite by confluence to form the lymphatic network, an auxiliary fluid-collecting system which ultimately conducts fluid back to the blood in the veins (Fig 3 5)

Distribution of Blood Volume, Pressure, Flow, and Resistance

The proportional distribution of blood volume in systemic arteries, capillaries, and veins is depicted in Figure 3 2C The arteries contain about 20, capillaries 5, and veins 75% The relative distribution between the pulmonic and systemic circulation is illustrated in Figure 3 6 Volume distribution estimates for an adult man appear in Table 3 2 and for the mesenteric circuit of the dog in Table 3.3

Pressure gradients typical of the human circulatory system are diagrammed in Figure 3 2C and 3 7, volume flow in Figure 3 2A, and blood flow velocity in Figure 3 2B A tabulation of blood velocity in the systemic circuit is given in Table 3 4

Chief among the hemodynamic features of the cardiovascular system are those determined by the mechanical relations between the pressure flow gradients, blood volume distribution and capacity, and compliance and flow resistance of the vessels in different segments of the circulation In simplest terms flow, pressure, and resistance are related as follows

$$\text{Volume flow } (V) \quad \frac{\text{Pressure gradient } (\Delta P)}{\text{Resistance } (R)}$$

The equation $\dot{V} \quad \Delta P/R$ or $\Delta P = \dot{V}R$ is analogous to Ohm's law of electrical circuits $I = E/R$ or $E = IR$

Among the mechanical factors which determine resistance to flow are the length and total cross-sectional area of the conduits, i e ,

$$\text{Resistance} \propto \frac{\text{Length}}{\text{Cross-section area}} \quad \text{or} \quad R \propto \frac{L}{A}$$

Thus, widening of the vascular bed decreases resistance in segments of comparable length

The *equation of continuity* of flow states that velocity of flow (v) times the cross sectional area (A) of a vessel must be constant at all serial segments in the circuit, e g , in the simplest case, i e ,

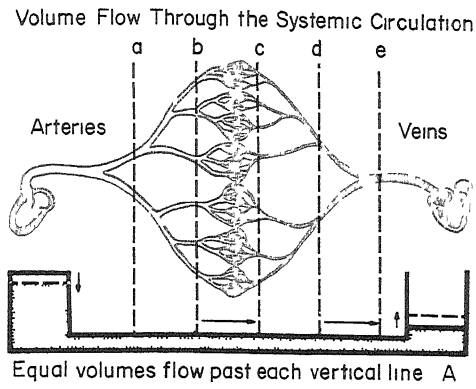
$$v_a \times A_a \quad v_b \times A_b, \text{ etc}$$

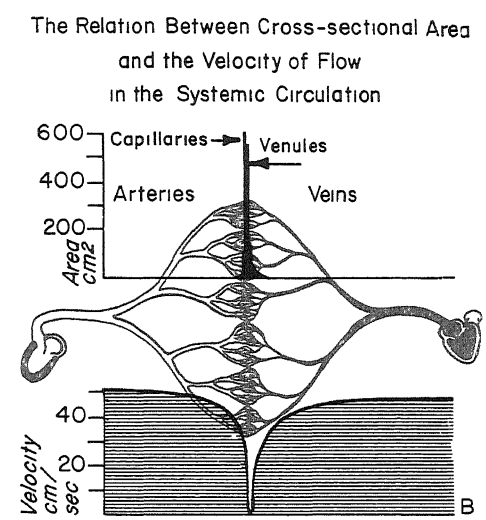
Thus, flow velocity is inversely proportional to total cross-section of the system,

$$v \alpha 1/A$$

These simple relations explain the characteristics of pressure and flow through the system depicted in Figures 3 2 and 3 7 In the complexly branching vascular system with pulsatile flow, these relations are not necessarily preserved at all sites and instants of time, but the concept holds as an appropriate generalization

Various functional types of vessels can be identified in accordance with pressure, flow, and volume distribution (see also p 3-164), e g high pressure vessels arteries, low pressure vessels veins, capillaries, pulmonary vessels, capacitance vessels veins, and resistance vessels. small arteries and arterioles chiefly,





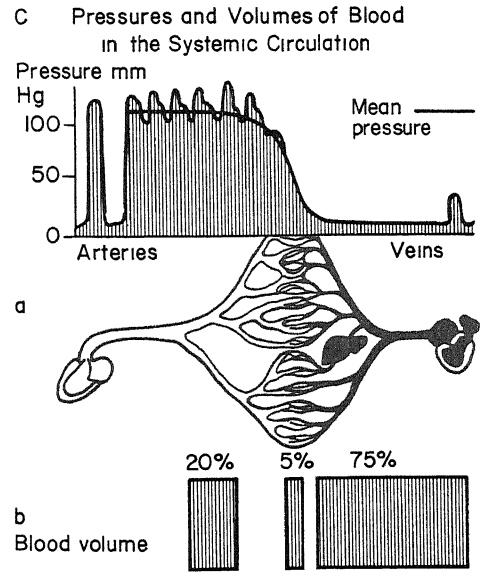


Fig 3 2 *A* The volume flow per unit time through each segment of the circulation must equal that entering and leaving the system despite great differences in cross sectional area *B* The cross-sectional areas of various segments of the circulation computed from a 13-kg dog are graphed above velocity of blood flow to show their inverse relationship Flow through the capillaries averages about 0 07 cm /sec *C* Approximate pressures and volumes throughout the circulation are depicted schematically Pressure is high and pulsatile in the arteries falls rapidly across the resistance vessels and the pulsation is damped out in the capillaries and veins Pressure is low and nonpulsatile in most capillaries and veins The blood volumes depicted are approximations (Redrawn from Rushmer 1961)

also capillaries and smallest venules Volume flow ($\dot{V}$) remains constant through each segment of the circulatory bed (Fig 3 2*A*) while mean pressure (P) decreases progressively (Figs 3 2*C* and 3 7) Pressure gradients (ΔP) are not uniform By scanning the pressure curves in Figures 3 2*C* and 3 7, the location of points of high resistance (R) can be found where pressure drop (ΔP) is greatest Resistance to flow is determined by the viscosity of the fluid and geometry of the tubular system Assuming viscosity to be constant, the factors determining resistance are length of conduits, their cross-sectional area and whether they are arranged in series or parallel In the case of dichotomously branching arterioles and capillaries, the resistances are in parallel and series The total resistance (R_T) at any segment is the sum of the vessel resistances in that segment The relative contributions can be derived from geometrical data, in Table 3 3, e g , in the case where areas and lengths are equal,

$$R_T \propto (1/\text{No of vessels}) \times (\text{Length/Cross-sectional area})$$

where the vessels are in parallel

Series resistances simply add, e g ,

$$R_T = R_{\text{arteriole}} + R_{\text{capillary}} + R_n$$

or the total resistance can be derived from

$$\dot{V} = \Delta P/R \text{ or } R = \Delta P/\dot{V}$$

While blood volume distributions (Figs 3 2*C* and 3 6, and Table 3 2) are not rigidly fixed, they are ordinarily maintained within fairly narrow limits The volume capacity of the systemic arterial distributing system (high pressure system) is the most limited and least variable, i e , it is least dilatable It contains (Table 3 2) only about 14% of the blood volume The

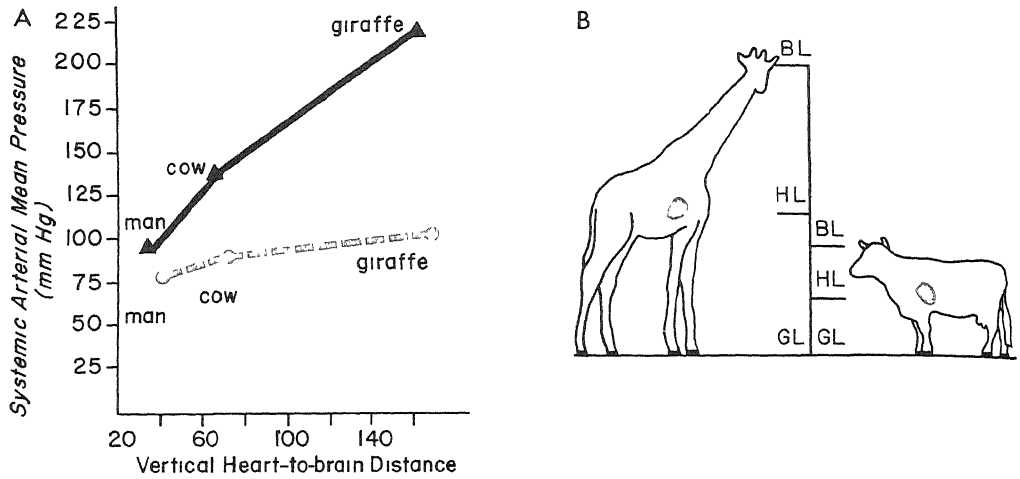


Fig 3 3 *A* mean arterial pressures at the heart level (*solid line*) and brain level (*dashed line*) in man cow and giraffe *B* scale drawing of giraffe and cow indicating relative levels of ground (G L) heart (H L) and brain (B L) (Redrawn from Patterson 1965)

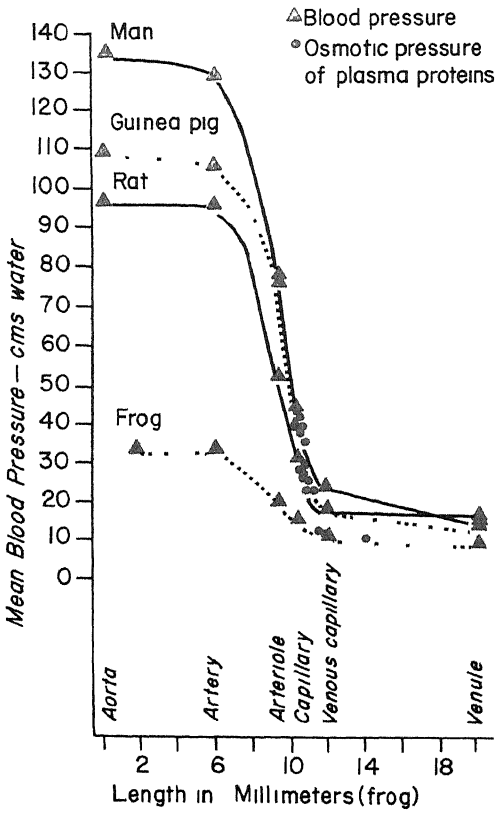


Fig 3 4 The relation between blood pressure gradients capillary blood pressure and osmotic pressure of the plasma proteins in four species (Redrawn from Landis, 1964)

Table 3 1
Relation of colloid osmotic pressure and blood pressure

Animal	Colloid Osmotic Pressure	Blood Pressure
	mm Hg	*mm Hg*
Aplysia	0 02	1 8
Homarus	1 15	8 5
Octopus	3 24	40
"Amphibia"	4 2	50
Oryctogalus	20 2	99
Homo	29 1	120

From Florey, 1966 Data from Florkin and Blum, 1934

remainder is in the heart (about 7%) and the low pressure system (about 79%) Of the latter about 9% is in the pulmonary circuit The low pressure system thus has the greatest capacity and is most dilatable When blood volume is increased as by transfusion, only about 1% is distributed to the high pressure system Thus, the low pressure system serves a reservoir function It is in the low pressure system that receptors for blood volume control are located which are the sensors for fluid volume regulations

The velocity flow (v) varies with the cross-sectional area of the system (Fig 3 2*B*) in accordance with the *equation of continuity* This provides for rapid flows through the relatively narrow distributing arterial and collecting ve-

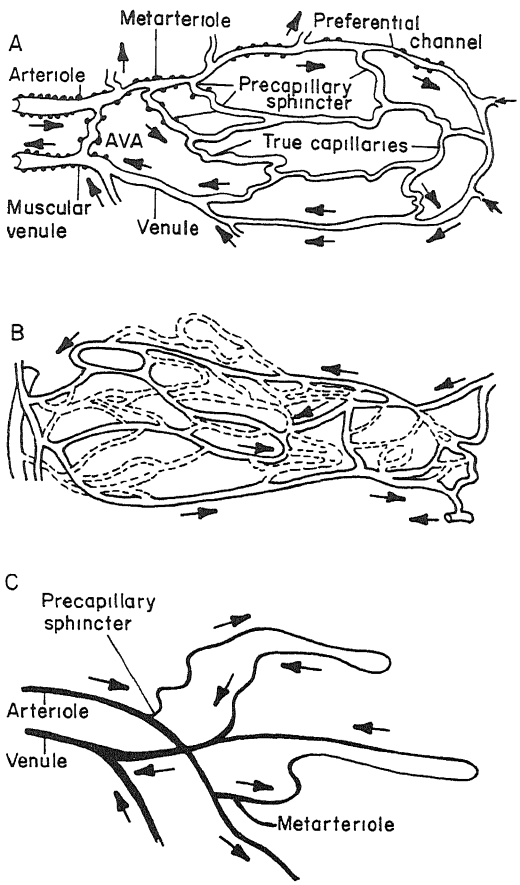


Fig 3 5 Three patterns of terminal vascular beds *A* muscle with preferential channels *B* mesentery true capillaries outlined by dotted lines *C* capillary loops of human nail bed (After Zweifach B W 1949 redrawn by Burton 1965)

nous networks, and slow movement at capillary diffusing and filtration (i e , exchange) sites

Regulatory Factors

Complex control arrangements maintain and adjust the distribution of blood pressure, flow, and volume to the various organs within physiological limits

INTRINSIC CONTROL

In the most peripheral part of the circulatory system, through the dimly understood mechanisms of autoregulation, flow is controlled so that with elevation of tissue metabolism flow increases, and vice versa Local cardiovascular control mechanisms govern the caliber of pe-

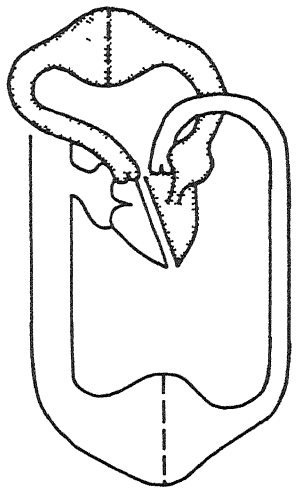


Fig 3 6 Circulation schema depicting blood volume distribution and distensibility of various segments of the circulatory system The shaded segment is the so-called central blood volume region It acts as a filling reservoir for the left ventricle with adjustable volume and compliance This accommodates temporary volume flow changes between the two ventricles and through reflex alterations in vascular compliance determines velocity of flow and pressure transmission across the pulmonary vascular bed In addition this low pressure segment of the circulation contains receptors which sense blood volume changes These initiate reflexes important in fluid volume regulation (From Gauer 1968)

Table 3 2

Estimated distribution of blood in vascular system of hypothetical adult man[a]

Region	Volume			
	ml		%	
Heart (diastole)	360		7 2	
Pulmonary				
Arteries	130		2 6	
Capillaries	110	440	2 2	8 8
Veins	200		4 0	
Systemic				
Aorta and large arteries	300		6 0	
Small arteries	400		8 0	
Capillaries	300	4,200	6 0	84 0
Small veins	2,300		46 0	
Large veins	900		18 0	
	5,000		100	

From Milnor, 1968

[a] Age, 40 years, weight, 75 kg , surface area, 1 85 m ^{2}

Table 3 3
Estimates of relative blood volumes in segments of the circulatory system from data on the mesenteric vascular bed of the dog

Structure	Diameter	Number	Total cross-sectional area	Length	Total volume
	mm		*cm²*	*cm*	*ml*
Aorta	10 0	1	0 8	40 0	30
Large arteries	3 0	40	3 0	20 0	60
Main arterial branches	1 0	600	5 0	10 0	50
Terminal branches	0 6	1,800	5 0	1 0	5
Arterioles	0 02	40,000 000	125 0	0 2	25
Capillaries	0 008	1,200 000,000	600 0	0 1	60
Venules	0 03	80,000 000	570 0	0 2	110
Terminal veins	1 5	1 800	30 0	1 0	30
Main venous branches	2 4	600	27 0	10 0	270
Large veins	6 0	40	11 0	20 0	220
Vena cava	12 5	1	1 2	40 0	50
Total					910

Data of F Mall Cited from Green et al , 1944

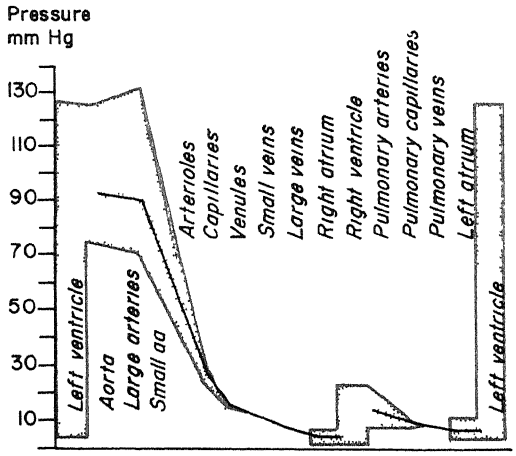


Fig 3 7 Pressure distribution in the human cardiovascular system The shaded area extends between diastolic and systolic pressures, indicating the amplitude of pulse pressure the central solid line represents mean pressure (From Milnor 1968)

ripheral arteries and thus influence cardiac output They have been shown to operate in the absence of central nervous system and peripheral autonomic innervation (Granger and Guyton, 1969) These controls provide a fundamental regulation which maintains homeostatic stability In individual organs autoregulatory responses occur within 1 to 3 min , while in the whole body their onset is delayed 3 min , or more and requires up to 50 min or more for completion Superimposed on this is reflex regulation (e g , baroreceptor reflexes) which is rapid (appearing within time units of less than 1 min) and long-term adjustments requiring days or weeks (e g , alterations in blood and other fluid compartment volumes)

The basic function of circulation being regulated is the filtration-diffusion process at the exchange vessels all regulations serve to maintain this at the appropriate level set by tissue metabolic needs

Accordingly, a logical place to begin in considering circulatory control is with the microcirculation, where fluid exchanges between blood and tissue fluids occur

Autoregulation Autoregulation is the term applied to blood flow changes in tissue and organs produced by local (or intrinsic) control mechanisms The organ or tissue tends to minimize the effect of changes in perfusion pressure on flow, thus holding flow relatively constant, when pressure is the only variable changed On the other hand, blood flow will increase with elevation of metabolic activity, even though the perfusion pressure remains constant These relations are schematized in Figure 3 8

Autoregulation has been demonstrated in the circulations of the brain, heart, kidney, liver, intestine, and muscle It probably occurs in most, if not all, tissues However, attempts to demonstrate autoregulation in the cutaneous or pulmonary circulations have failed These are called *passive vascular beds* In them there is little intrinsic tendency for flow to be kept constant when arterial pressure is altered, or to

Table 3 4[a]
Vessel diameter blood velocity and Reynolds number for the systemic circulation of man

Structure	Diameter	Blood Velocity	Tube Reynolds[b] Number
	cm	*cm /sec*	
Ascending aorta	2 0–3 2	63[c]	3600–5800
Descending aorta	1 6–2 0	27[c]	1200–1500
Large arteries	0 2–0 6	20–50[c]	110–850
Capillaries	0 0005–0 001	0 05–0 1[d]	0 0007–0 003
Large veins	0 5–1 0	15–20[d]	210–570
Venae cavae	2 0	11–16[d]	630–900

[a] From Whitmore, 1968
[b] Assuming viscosity of blood is 0 035 poise
[c] Mean peak value
[d] Mean velocity over indefinite period of time

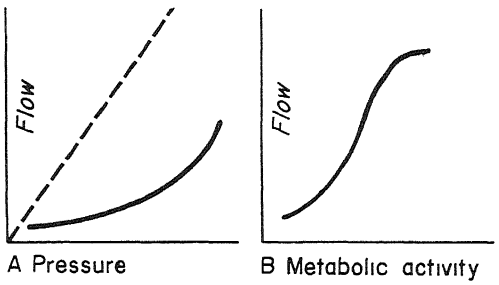


Fig 3 8 *A* The broken line depicts a linear relationship between pressure and flow as would be seen when a Newtonion fluid perfuses a rigid tube system The solid line depicts the pressure-flow relationships when a tissue with constant metabolic activity is perfused at increasing pressures *B* A hypothetical relationship between the level of metabolic activity and flow at constant perfusion pressure is depicted

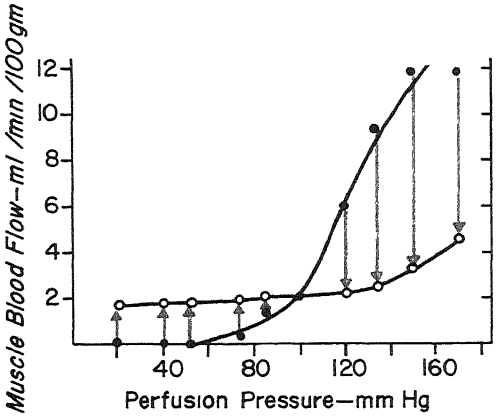


Fig 3 9 Pressure-flow relationships in the skeletal muscle vascular bed of the dog The closed circles represent the flow rates first measured when pressure was changed abruptly from the control level of 100 mm Hg the open circles are the steady-state flows finally attained at the new perfusion pressure The two lines intersect at the control pressure point (From Berne and Levy 1967)

maintain any other variable in the face of tissue metabolic changes

In Figure 3 9, pressure-flow relations are graphed from observations on an isolated skeletal muscle preparation in the resting state Note that the immediate effect of increase in pressure was elevation of flow, which then after 30 to 60 sec , returned toward the control level

Since autoregulation occurs in denervated tissues, reflexes are not necessarily involved However, in the intact animal reflex and hormonal control mechanisms may override local regulation in a particular organ. Several factors have been suggested to account for local circulatory control Among these are. contractile response of vascular smooth muscle to increased tension, variations in amounts of vasodilator metabolites, variations in oxygen tension, and changes in tissue pressure Other possible mechanisms are discussed on p 3–168 In any case, autoregulation is clearly an important local regulatory factor accounting for blood flow distribution consonant with tissue needs Through the mechanisms of autoregulation, nutrient flow and capillary pressure are regulated within physiological limits

In experimental dog preparations in which all nervous control of circulation has been removed by destroying the central nervous system, the vascular system as a whole exhibits substantial, slowly developing autoregulatory responses over periods of 20 to 90 min These responses could develop further with time (Granger and Guyton, 1969) This finding suggests that whole body autoregulation is a deliberately acting control mechanism which ranks along with ner-

vous and humoral control of cardiac output and peripheral resistance in regulating blood pressure

EXTRINSIC CONTROL

Local adjustments of blood flow by intrinsic regulation depend on maintenance of adequate blood pressure within the distributing arteries The mechanical relationships of flow, pressure and resistance in the systemic arterial circulation are, as given before

$$\dot{V}_{as} \quad \Delta P_{as}/R_{as}$$

where $\dot{V}_{as}$ is systemic arterial flow, ΔP_{as}, systemic arterial pressure gradient, and R_{as}, systemic arterial resistance It is clear that when vessels dilate locally and resistance (R_{as}) is lowered, adjustments must occur in flow ($\dot{V}_{as}$) and pressure (P_{as}) All tissues require that a certain pressure be maintained for their autoregulatory mechanism to be effective This is provided for by extrinsic regulatory systems which govern heart rate, blood vessel caliber, and blood volume The components of these extrinsic control systems can be classified as follows effectors heart muscle, vascular smooth muscle, endocrine secretory cells, regulators nerve impulses, hormones, vasoactive substances, sensors nerve receptors, and transmitters nerves, blood stream

A simplified mechanical analog for certain aspects of blood pressure control is the fluid tank with fluid level control (Fig 3 10) The fluid level variable (h, height) is kept constant by a mechanical link, L, which adjusts a valve, a, in accordance with the effect of pressure (height × gravitational acceleration, hg) on an expansible syphon, S The valve, a, regulates flow in, $\dot{V}_i$ Flow out, $\dot{V}_o$, is determined by the fluid level, h

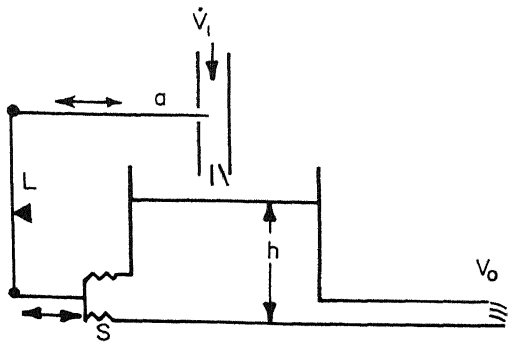


Fig 3 10 Tank with level control

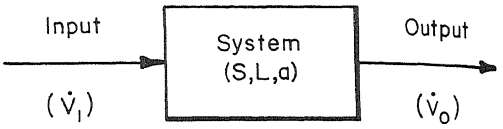


Fig 3 11 Input-output system

This is an *input-output system* with negative *feedback control* An increase in pressure (hg) leads to a decrease in inflow, $\dot{V}_i$

In this system, input and output will be related in accordance with the characteristics of the negative feedback mechanical components S, L, and a (Fig 3 11)

Often, in biology the components of control systems and their interactions are only vaguely understood As a first approximation it is useful to study merely the relationship between the input and output of a given variable, without analyzing the components of the system responsible for the relationship Such a relationship between input and output is called a *transfer function* (Fig 3 12) In some cases the transfer function can be represented by a mathematical equation, in others, only by a table of observed values or a set of postulates or rules [1]

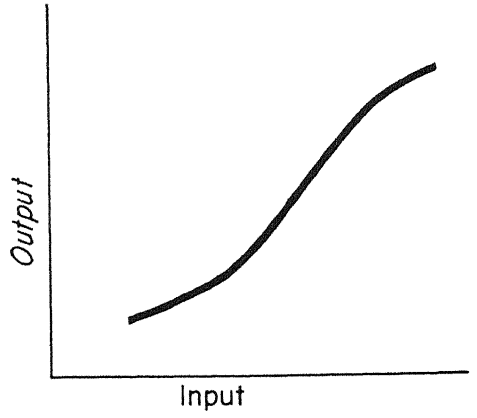


Fig 3 12 Transfer function

The regulation of circulation has as its chief function the maintenance of an adequate *flow* of blood to all organs Given a constant pressure in the distributing arteries, flow can be altered to suit tissue metabolic needs by local

[1] The term 'transfer function' is not used here and subsequently in accordance with its mathematical definition in which for any linear component or any linear system it designates the 'Laplace transform of the output divided by the Laplace transform of the input with zero initial conditions assumed (Riggs, 1970)

adjustment of peripheral resistance However, an increase in flow to one organ would tend to decrease flow to another insofar as a fall in pressure were allowed to occur This is prevented by *sensors* which monitor blood pressure These are mechanoreceptors (stretch receptors) which are activated by changes in arterial pressure and are thus known as *pressoreceptors* or *baroreceptors* Located in the walls of blood vessels (e g , carotid sinus, aortic arch), they indirectly help govern blood flow by detecting changes in blood pressure These are the sensory elements of a negative feedback loop which helps maintain the constancy of blood pressure The *effector* organs are the heart and peripheral arterioles whose level of activity is regulated by *transmitters*, nerve impulses, and humoral agents Thus there is a *controlling system* made up of *regulators* (nervous system, endocrine organs) and a *controlled system* (heart and blood vessels) which govern a *regulated variable* (blood pressure)

These elements of the cardiovascular control system can be formulated as a block diagram in which the regulators are designated *controlling system* and the effectors, *controlled system* (Fig 3 13)

In such a system, certain elements can be examined separately to establish their relations (*transfer functions*) For example, one of the ways the baroreceptors affect blood pressure is through altering heart rate, which, in turn, increases or decreases cardiac output ($\dot{V}$) Thus, heart rate (frequency, F) is related to blood pressure in a baroreceptor (e g , carotid sinus pressure, P_{cs}) somewhat as in Figure 3 14

In curve A, heart rate is plotted as a function (dependent variable) of carotid sinus pressure Represented in curve B is the fact that heart

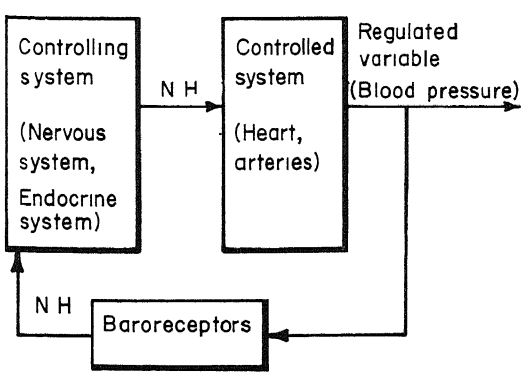


Fig 3 13 Block diagram of a cardiovascular control system *N* nerve impulses *H* humoral agents

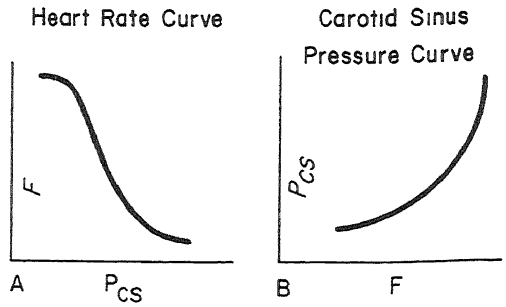


Fig 3 14 Relation of heart rate F and carotid sinus pressure P_{cs} (These and other similar curves are drawn to show the sign of the slope their shape has no mathematical validity)

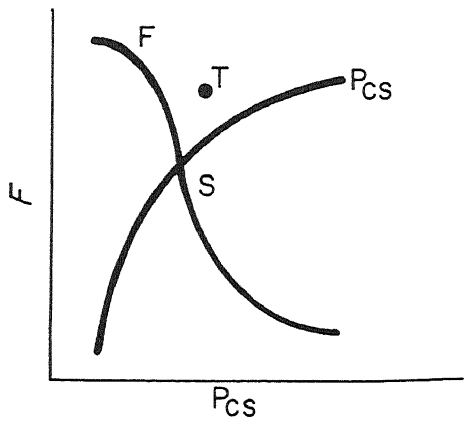


Fig 3 15 Determination of equilibrium (steady-state *S*) values of heart rate F and carotid sinus pressure P_{cs}

rate, as one determinant of blood pressure influences carotid sinus pressure in a predictable way Note that in each graph of a transfer function the input (independent) variable appears on the abscissa, the output (dependent) variable on the ordinate These two graphs illustrate interdependent variables, a common situation in biological control systems How these relations interact to help maintain the blood pressure within certain limits can be appreciated by plotting them on the same graph In this plot the graph B for carotid sinus pressure as the dependent variable is rotated 90° and the carotid sinus pressure ordinate reversed The resulting graph shows an intersection, S, of the curves in A and B (Fig 3 15)

The point of intersection of the two transfer function curves represents the steady-state heart rate and carotid sinus pressure when the system is at rest or undisturbed If some transient influence (e g , exercise, excitement) dis-

places the heart rate and carotid sinus pressure curve point of intersection, say to point T, the two curves will return toward their original positions until they intersect at S, the resting steady-state position or point of equilibrium

The transfer functions depicted in Figures 3 14 and 3 15 show only the influence of carotid sinus pressure on heart rate and heart rate on carotid sinus pressure However, both these variables are subject to other influences These influences can alter the transfer function of either dependent variable Suppose that the sensitivity of the receptor nerve terminals in the carotid sinus were increased by a drug Then its transfer function would be shifted to the right as indicated by the dotted line in Figure 3 16

Thus, for a given carotid sinus pressure, the heart rate would be lower and the equilibrium point would be shifted to the right, a condition which would permit the blood pressure at rest to remain lower than before administration of the drug Should some factor inhibit the heart rate, its curve would shift to the left and the point of intersection change accordingly (Fig 3 17)

A slightly more complex example of control can be illustrated by diagramming two factors known to modify the pumping action of the left side of the heart The basic model is that of a three variable system in which two independent variables, x and y, alter dependent variable z (Fig 3 18)

Cardiac output (left ventricle volume flow, $\dot{V}_{LV}$) is a product of the amount of blood ex-

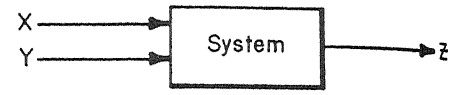


Fig 3 18 A three variable system

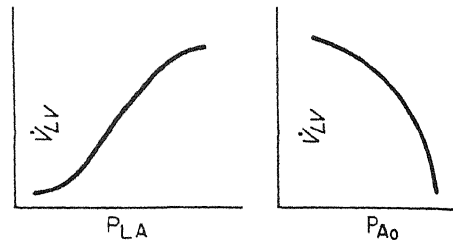


Fig 3 19 The influence of left atrial pressure P_{LA} and aortic pressure P_{Ao} on left ventricular output $\dot{V}_{LV}$

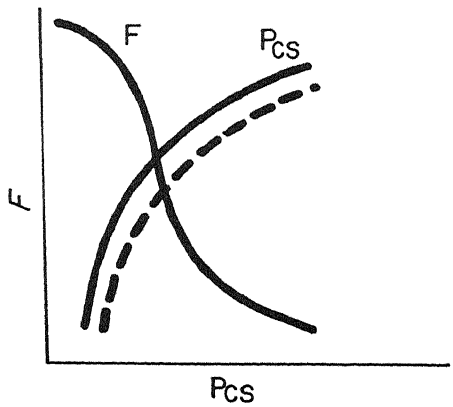


Fig 3 16 The effect of increasing sensitivity of the carotid sinus pressure receptor F heart rate P_{CS} carotid sinus pressure

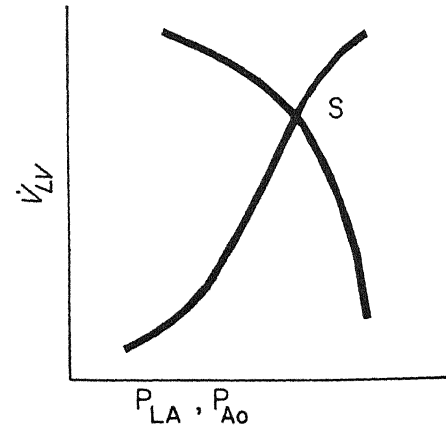


Fig 3 20 Equilibrium or steady-state point S for the combined effects of left atrial pressure P_{LA} and aortic pressure P_{Ao} on cardiac output, $\dot{V}_{LV}$

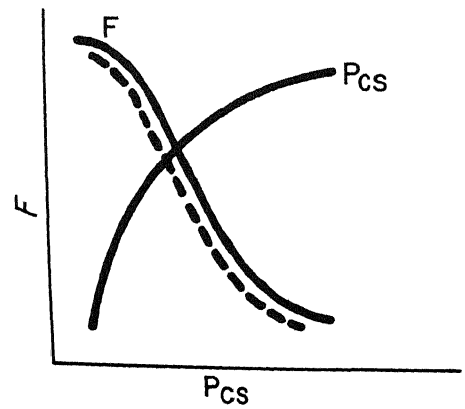


Fig 3.17. The effect of inhibition of the heart rate on the relationship between carotid sinus pressure P_{CS} and heart rate, F.

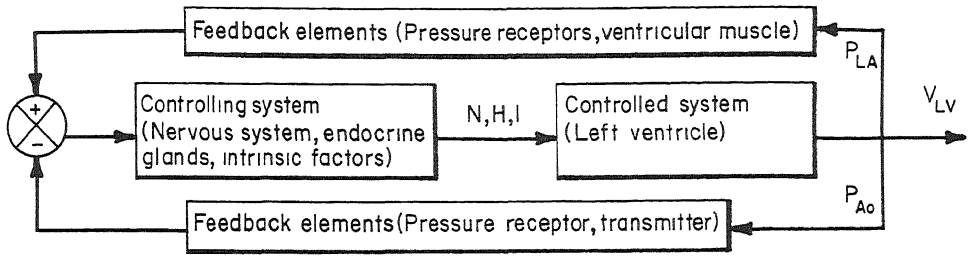


Fig 3 21 Block diagram of a control system regulating left ventricular output V_{LV} through two simultaneously acting feedback loops which converge P_{LA} atrial pressure P_{Ao} aortic pressure N nerve impulses H hormones I intrinsic ventricular muscle factors ⊕ convergence point a symbol representing the interaction of the two feedback loops which in this case have opposing effects

pelled with each heart beat (stroke volume, Vs), times the heart rate (frequency, F)

$$\dot{V}_{LV} = F\ Vs$$

Two independent variables which modify left ventricular output are the input pressure in the left atrium (P_{LA}) and the output pressure load (aortic pressure, P_{Ao}) These are related by transfer functions as in Figure 3 19

The intersection of these two curves represents a steady-state condition, s, for given values of P_{LA} and P_{Ao} which determine $\dot{V}_{LV}$ (Fig 3 20)

The two independent variables, P_{LA} and P_{Ao} can be diagrammed as a three variable system with two feedback loops, which act upon a controlling system (nerves, hormones, intrinsic ventricular muscle factors) to regulate a controlled system (left ventricle) such that its output tends to remain stable (Fig 3 21)

By experimentally holding one of these independent variables constant and altering the other, their individual roles in governing $\dot{V}_{LV}$ can be determined Results of such experiments in dogs are shown with a three-dimensional plot in Figure 3 22 In this figure it can be seen that at low left atrial pressures (P_{LA}) an increase in aortic pressure (P_{Ao}) has little effect on $\dot{V}_{LV}$ until high aortic pressures are attained. At high left atrial pressures, the influence of increased aortic pressures is greater Similarly, at high aortic pressures, decreases in left atrial pressures have little effect until low atrial pressures are reached However, at low aortic pressures, the influence of increasing left atrial pressure (P_{LA}) on $\dot{V}_{LV}$ is much greater

Thus far only control arrangements governing cardiac action have been considered Resistance (R_{as}) to flow out of the arterial system ($\dot{V}_o$) is also regulated in a similar way to

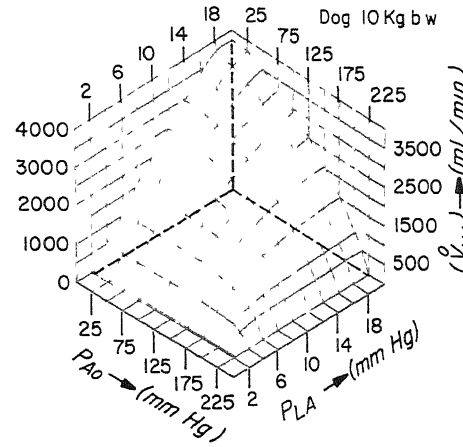


Fig 3 22 A three dimensional plot of mean aortic pressure (P_{Ao}) mean left atrial pressure (P_{LA}) and cardiac output ($\dot{V}_{LV}$) (From Milhorn 1966 by K Sagawa)

oppose changes in blood pressure Constriction and dilatation of arterioles, causing alterations in resistance, occur in concert with cardiac output changes to maintain homeostasis of blood pressure Intrinsic factors within the tissues adjust arteriolar caliber to permit flow changes appropriate to tissue metabolic needs Feedback control loops involving the baroreceptor sensors also act through nerves and hormones to produce vasoconstriction or vasodilatation in various tissues, as required to maintain blood pressure and adequate flow

Although the chief requirement of these control systems is to maintain an adequate *flow* of blood, the main sensors detect pressure changes Flow receptors are not known in the circulatory system

The controls described thus far act rapidly to effect their adjustments in a short period of time i e , *short-term* regulation Such adjust-

ments are effected within time intervals of seconds, minutes, or, at most, a few hours

Long-term regulation is also present, effected through mechanisms which govern body fluid volume, vascular tone, and vascular capacity These in turn alter blood pressure, cardiac output, and vascular resistance, producing or opposing sustained deviations in pressure and flow distribution Long-term adjustments require days or weeks and thus operate slowly

For further reading see Adolph (1968), Bayliss (1966), Engelberg (1966) Granger and Guyton (1969) Yamamato and Brobeck (1965), Johnson (1964) Milhorn (1966, 1969), Reeve and Guyton (1967) Riggs (1963, 1970), Landis (1964), Patterson et al (1965), Florey (1966) Burton (1965) Gauer (1968), Milnor (1968), Green et al (1944), Whitmore (1968) and Berne and Levy (1967)

Functional Characteristics of the Systemic Circulation

VOLUME FLOW

The anatomical complexity of the circulation may obscure some basic facts which have been stressed The volume flow of blood through any part of the circulation must equal that flowing through any other part There may be regional shifts of flow so that some arteries or capillaries carry more or less flow than in other arteries or capillaries, or there may be a small difference in the blood volumes put out by the two ventricles, but the former does not affect the relative total flow through the capillaries and arteries while the latter can only exist momentarily for a few heart beats, else the organism dies

BLOOD CONTENT

The blood content of different areas is not constant but is continuously changing and is available for redistribution to other vascular areas where it may be needed The volume in any one compartment depends not only upon the basic anatomical size of the bed but upon dynamic changes induced by passive and active changes in capacity and content, i e , distensibility

Basic control. Blood storage can be on a long term basis (blood reservoir) or it can change rapidly during a cardiac cycle (compression chamber). Both are dependent on the blood content of the cardiac cavities and of the vessels at zero internal pressure and on their distensibility, i e , change of diameter and length with change of internal pressure, in turn, the basic distensibility of the cavities is governed by the myocardial resistance to stretch For example, the distensibility of a heart cavity actively changes if, for the same resting size or volume, the pressure within the cavity is altered The distensibility of the ventricles can be grossly altered by nervous, humoral, and other influences The distensibility of the vessels is governed by their content of elastic and collagen tissue and smooth muscle (Fig 3 23) All varieties of blood vessels except the arterioles, capillaries, and venules, have a prominent component of elastic tissue, and in some vessels such as the aorta, as much as 40% by weight is elastic tissue The elastic fibers with their great range of extensibility before their elastic limit is reached, produce maintenance tension without energy expenditure against the normal blood pressure In the wall of every type of blood vessel except the capillary, there are white collagenous fibers in a matrix penetrating between the other elements The collagenous fibers act in a similar way to the elastic fibers but, because of the architecture of the wall, are stretched only at higher than normal blood pressure and have a protective supporting role (Fig 3 23) If, however, the vessels reacted in a purely passive way based on their content of collagen and elastic fibers, their internal pressure would be determined only by the volume of contained blood and, at any constant mean pressure the mean volume of blood within them would also not change However, many types of vessel contain much smooth muscle which does not contribute significantly to the elastic tension of the wall Its function is to produce active tension under vasomotor stimulation or local action of various chemicals (vasodilator and vasoconstrictor substances and oxygen lack) and so change the size of the lumen Such vascular tone adjustment is seen in most vascular areas It counteracts distensibility by inducing smooth muscle reaction which, at times, is strong enough to induce a net decrease in vascular lumina in the presence of an elevated pressure Such activation will, of course, change completely the pressure-volume relationship in a vascular bed or area

Aorta and its Arterial Branches. The pressure-volume relationship or distensibility of the isolated aorta has been exhaustively studied in vitro The distensibility varies widely in different aortas in the same age group (Fig. 3 24) There is no statistically significant correlation

between age and absolute distensibility However, aortic size or capacity increases considerably with age so that the average diameter of the aorta may exceed that of the sum of its branches and the contained volume at a diastolic pressure of 80 mm Hg may almost double This shift toward larger volumes with increasing age indicates that the enlarged aorta does not now need to expand as much to accomodate the same systolic ejection The distensibility of an isolated aortic segment studied in vivo and in vitro differs mildly during its inflation and deflation, the resulting curves forming a loop (Fig 3 25) This is presumably because of hysteresis of the elastic fibers

Although such studies aid in orientation, there is no guarantee that the characteristics of the aorta or, indeed, any vessel postmortem apply to conditions in the intact normal animal or man Since the systemic arterial system contains only about 14% of the total blood in the systemic circulation, even if smooth muscle contraction reduced its volume by 25% (or 175 ml) its function as a reservoir would be small Actually, during systole and cardiac ejection, much of the pressure and flow are transferred immediately to the periphery However, the aorta and some of its immediate branches belong to the elastic type of arteries, i e , elastic tissue is abundant and muscle fibers are relatively scarce Because of their large distensibility, the walls of these large central arteries act as compression chambers which serve to buffer the more peripheral branches from too sudden an increase in pressure and flow The volume of blood suddenly injected into the aorta by the powerful ventricular action is only partly moved forward, a considerable portion is stored locally by the successive expansion of the large vessels Although earlier work indicated that there is no radial expansion of the aorta beyond the arch, later work using cinematographic

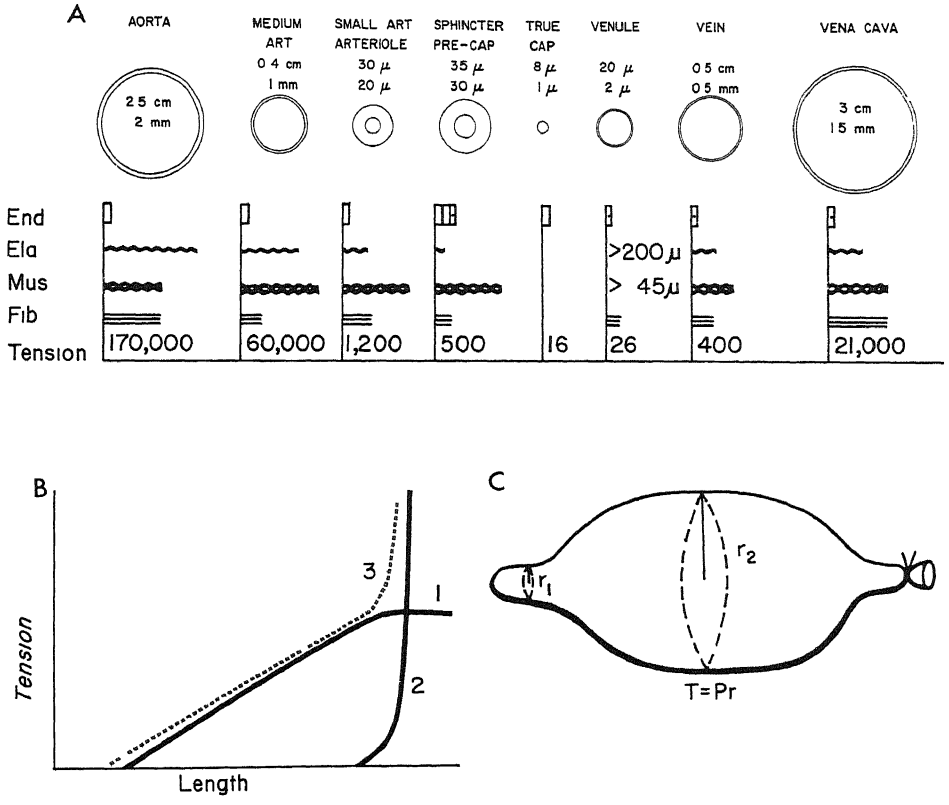


Fig 3 23 Illustrated are the lumen diameters wall thickness relative amounts of different tissues and wall tensions in the various blood vessels *Lower left* Tension-length diagrams of elastic *1* and collagen fibers *2*) presented alone and in combination *3*) *Lower right* balloon demonstrating pressure-tension relation The pressure (*P*) within the balloon is equally transmitted to all parts of the contained air (Pascal s law) and the tension (*T*) in the balloon varies with the radius (*r*) in that portion Laplace s Law (Modified after Burton 1954 and Wolf 1952)

films at 1500 frames/sec showed dilation of the thoracic aorta during systole of 15 to 20% of its diastolic diameter which corresponds to a 30 to 40% increase in volume per unit length of the artery, radial enlargement of the more distal aorta was somewhat less. Thus it is estimated that up to one-half of the pressure and flow discharged during systole is stored here and that this moves forward during diastole through the elastic recoil of the arterial walls. This enables a portion of the blood to run off into the capillaries during diastole and not during systole.

Whether active changes in aortic distensibility normally occur is only incompletely investigated. Simultaneous measurements have been made of phasic aortic blood pressure by an appropriate pressure manometer and of changing aortic circumference during a cardiac cycle. In the open-chest dog or man the latter is obtained by noting the movement of a pair of recording mechanical calipers placed around the aorta. In the unanesthetized dog, aortic circumference is detected by measuring the changing resistance in a mercury-filled rubber tube wrapped around it. In the open-chest dog, when the arterial blood pressure is elevated to a high level by mechanical constriction of the aorta, the size of the aorta during diastole initially increases but then decreases in the presence of an elevated internal pressure, and at the same time, its distensibility increases, i.e., the pulse pressure becomes larger. A large intra-arterial injection of adrenaline constricts an aortic segment being perfused in vivo so that it may contain less than half the volume of the normal aorta in the low pressure range, while in the high pressure range the volumes are almost equal, i.e., the constricted aorta now has a much greater distensibility than the normal aorta. These changes are presumably related to active changes in the smooth muscle which is abundant in these vessel walls. However, these experiments must be interpreted cautiously, for preliminary measurements in intact unanesthetized dogs have shown an intimate relation between the aortic pressure and aortic circumference. Thus, the actual amount of and the significance of active alteration of aortic size under changing physiological states is not known.

The more peripheral arteries such as the carotid, brachial, and femoral and their branches, are predominantly muscular types of arteries

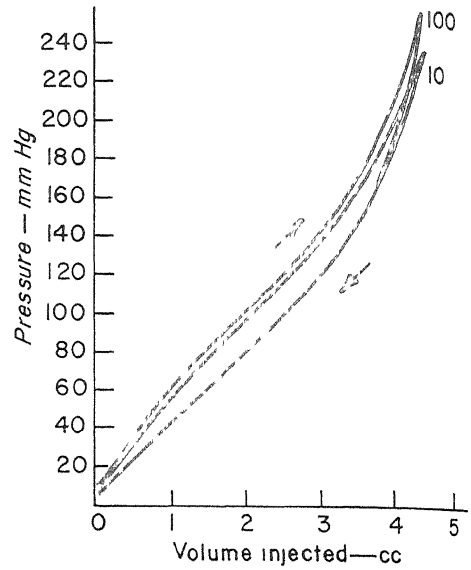


Fig. 3.25. Pressure volume diagrams obtained from an isolated segment of thoracic aorta by injections and withdrawals (*arrows*) at rate of 10 and 100 ml of blood per minute (Alexander 1954a)

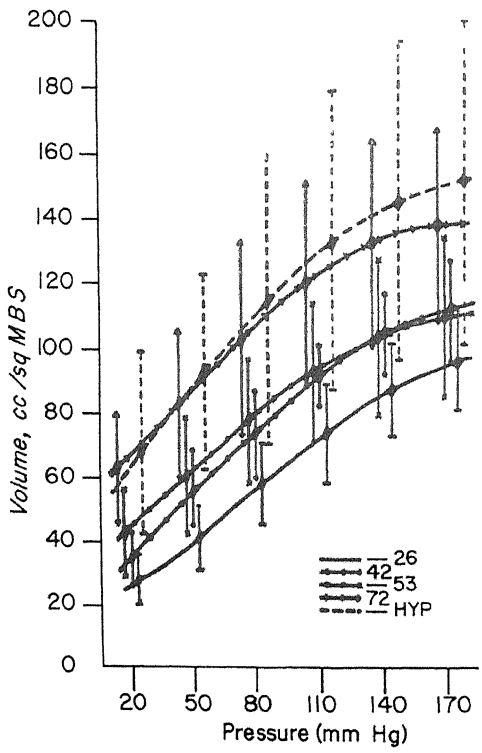


Fig. 3.24. Influence of age on the volume-pressure relations of human aortas. Each curve represents the average values for a given age group. Vertical lines represent standard deviations (Remington and associates 1948)

containing circularly arranged smooth muscle fibers in large numbers which, by their shortening, can reduce the lumen of these vessels considerably Such a set-up obviously permits adaptation or adjustment of this distributing system to accommodate changing volumes of circulating blood This is initiated by heat, cold, trauma, and many other physiological events In the latter instance, pressure-volume curves done in situ on small (1 2 mm or less internal diameter) and very muscular arterial segments of the intestine showed that with elevation of the aortic distending pressure, directly or through an increased portal venous pressure, the segment diameter may actually decrease Since intra-arterial procaine did not block the response and since the vessel constriction was abolished after ischemia, papaverine or cyanide, an active response of the arterial smooth muscle wall (and probably of myogenic origin), is suggested which is caused by the elevated wall tension

It can also be demonstrated that the femoral artery dilates in response to contraction of the peripheral muscles This persists with local nerve section, occurs only on the stimulated (active muscle) side, and is abolished by local procaine or cutting of the artery distal to the point of measurement of femoral artery volume Therefore, the response is by way of a conducting system in the arterial wall The velocity of such peripheral conduction is 10 cm/sec , and therefore, the conducting elements are probably the smooth muscle of the media of the vessel wall

Arterioles The arterioles contain the smallest blood volume of the cardiovascular system, but the pressure and flow in the circulatory system are more sensitive to minute changes in blood content of the arterioles than to that of any other part of the circulatory system The arterioles are the final or end branches of the distributing system and operate as stopcocks which control the run-off of blood from the arterial system into the capillaries The caliber of these vessels is controlled by the state of contraction of their circularly disposed and strongly developed smooth muscle fibers The latter are regulated by local metabolic process and by blood borne substances They are also supplied with vasomotor fibers which are controlled by centers in the spinal cord and medulla Their distensibility other than through myogenic change has not been determined.

Capillaries The blood content of the capillaries (300 ml) is also very small The arteries, even as far as the commencement of the true capillaries, are impervious The total filtering surface by the capillary endothelium in the adult human body amounts to about 6300 m ^{2} (68,000 ft ^{2}) The summed areas of the capillary walls in all regions could be imagined, therefore, as constituting an endothelial membrane over 12 miles long and a foot wide, yet so thin that when tightly rolled, it would form a cylinder of about the thickness of a lead pencil

The capillary wall proper is composed of a single layer of endothelial cells and is about the most tenuous structure conceivable (Fig 3 26) It does not exceed half a micron in thickness The endothelial cells which are flat are joined at their fringes by a so-called "intercellular cement" which fills in between adjacent cells The cellular area makes up more than 99% of the total area across which exchange takes place, the small nonliving zone (cement) less than 1% The cement substance which binds the endothelial cells at their edges is being continuously washed away and renewed again by the endothelial cells Its basic structure is a porous network of a complex cement substance, presumably a calcium proteinate At one time it was thought to be hyaluronic acid, but this idea is no longer held Superimposed on this network, either by electrical, chemical, or surface tension forces, is a large adsorbed molecular component of protein which lines the inner surface and plugs most of the large pores in the cement, considerably reducing the perviousness of the structure The capillary endothelial tube is enclosed and supported by a delicate sheath or membrane of fine fibrils derived from or closely related to the surrounding connective tissue The intervening space between capillary and this membrane contains a freely moving fluid Surrounding this in the interstitial spaces

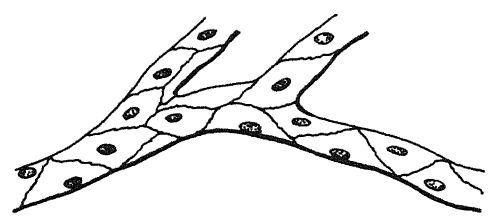


Fig 3 26 Section of capillary wall stained with silver nitrate to demonstrate the intercellular cement Nuclei of the cells stained with hematoxylin

is the so-called tissue fluid which is of a gel-like consistency

It is still debatable whether the bulk of exchange occurs across the entire surface area or is restricted for most molecules to the pervious intercellular cement Ions and small molecules diffuse across the capillary wall at such an incredibly fast rate (at least 100% exchange within 1 min) that it has been suggested that the whole capillary wall is permeable to these substances The penetration, however of capillary walls by water and dissolved substances appears to take place solely by processes which require no energy transformation by the capillary endothelial cells From this has arisen a general belief that the capillary walls, or specifically the intercellular cement, is pierced with numerous ultramicroscopic openings or pores with an average diameter of about 30 Å which are generally too small to allow the passage of plasma protein molecules, but are of sufficient size and number to account for the observed rates of passage of water and nonprotein constituents of the plasma Considerable evidence has been adduced that lipid soluble molecules, such as oxygen and carbon dioxide, can penetrate rapidly the plasma membranes of the capillary endothelial cells and are not restricted to the aqueous intercellular spaces Finally, the evidence is fairly clear that penetration of large molecular aggregates and formed elements, such as the red and white cells, occurs through the intercellular portion of the wall The presence of vesicles in capillary endothelial cells which can take up tagged protein molecules, suggests that a process analogous to pinocytosis plays a role in the transport of large molecules across the endothelium Capillary permeability is by no means uniform throughout the body For example, skeletal muscle capillaries are about 100-fold less permeable to water than the glomerular capillaries

The functional anatomy of the capillaries has been worked out for some cold-blooded animals, for the rat and dog mesentery through microscopic observations of the in vivo vessels, and for the myocardium by both anatomical and microscopic observations (Fig 3 5) Briefly, blood flows generally from the arterioles directly into a metarteriole, and then into capillaries The metarterioles lead directly into channels which are main thoroughfares from the capillary bed to the venules The true capillaries concerned with interchange between blood and tissues are interanastomosing side branches of the main channels through the bed At the ostia of each capillary is a small precapillary sphincter of smooth muscle which is controlled by nerves presumably from the sympathetic nervous system, in the same manner that these nerves control the arterioles and metarterioles In the body, the metarterioles and their precapillary sphincters undergo periodic contractions at intervals of 15 sec to 3 min When the tissue is in a resting state, the constrictor phase of this rhythm predominates and the precapillary sphincters may be completely closed When the tissue becomes active, the dilator phase of the metarterioles predominates and the precapillary sphincters are open Thus, in skeletal muscle, it is believed that the increase in blood flow with exercise comes in large measure from this opening up of large numbers of additional capillaries It is believed that the factors that affect the degree of constriction and relaxation of the metarterioles and the precapillary sphincters during vasomotion are the same factors that affect arteriolar diameters These are the nervous and hormonal stimuli and local conditions in the tissues Thus, sympathetic stimuli and epinephrine in the blood intensify the constrictor phase of vasomotion in most areas of the body in the same manner that these two constrict the arterioles On the other hand, vasodilator substances decrease the vasoconstrictor phase The significance and value of this arrangement generally to capillary beds and to overall circulatory dynamics remains to be established

Although controversy has raged for many years, at present it is believed that change in capillary diameter is on a purely passive basis, i e , by change in intracapillary pressure or in extravascular tissue pressure, supposedly, the active changes in capacity previously described by many workers is probably due to the activation of smooth muscle whose presence was unsuspected at that time in the metarterioles and precapillary sphincters

Capillary Exchanges. The hydrostatic and osmotic forces causing fluid exchanges across the capillary membrane are discussed on p 3-148, and in chapter 12 of section 4 They account for the net exchange, the "bulk flow", in which noncolloidal plasma constituents move in and out of the capillary together The concept, known in classical forms as the Starling hypothesis, supposes that filtration occurs

at the arteriolar ends of capillaries and absorption at the venular ends While this relationship may occur in a given capillary, a decrease in precapillary resistance may increase pressure and flow so that filtration takes place along the entire length of a given capillary Conversely, the capillary pressure may be reduced below that of colloid osmotic pressure so that the vessel acts entirely as an absorbing unit

Such bulk flow explains fluid shifts between the plasma and extracellular compartments However, it contributes little to the actual rate of exchange of materials between plasma and interstitial fluid Diffusion is the prime determinant of transcapillary exchanges and for a given solute may proceed along a concentration gradient in a direction opposite to net bulk flow

Veins Since the systemic veins contain such a large proportion of total blood volume, they must function as blood reservoirs The responses by which they may counteract distending forces and thus prevent excessive pooling of blood have been of great interest and have been documented in animals and human experiments As the venules merge to form larger veins, they acquire a layer of connective tissue and then distinct muscle fibers These vessels have a larger blood content than the corresponding arteries Their walls are also much thinner and contain less elastic tissue but have a rather well-developed muscular coat When the volume of blood in a normally partially collapsed vein is augmented, the intravascular pressure barely increases until the vein becomes round and has a certain degree of fullness The volume increase up to this point is similar to the filling of a collapsible tube without distending the walls True distension now occurs when further volume increments lead to an elevation of intravascular pressure (Fig 3 27) Veins excised or within the organism have a very low distensibility when the increase in volume is rather rapid With prolonged elevation of intravenous pressure, there is a delayed compliance or passive yielding of the venous walls which is attributed to the viscoelastic properties of the smooth musculature Thus, although initially the low distensibility of veins may prevent blood pooling, the delayed compliance favors it It is estimated that the ability of the venous system to take up blood is at least 100 times that of the whole circulatory system Thus, if 1000 ml of blood is transfused in a human, about 990 ml will reside in the extra-arterial or low pressure system, and 10 ml in the arterial system

The diameter of these venous capacitance vessels may be changed either passively by external forces (gravity, respiration, or pressure breathing) or actively by contraction or relaxation of their walls There is considerable evidence to indicate that the venous system (including the pulmonary system) behaves somewhat like an elastic bag For example, change in posture has no particular effect on their state of contraction when human subjects are bled or transfused with about 500 ml of blood the pressures in the central veins, the pulmonary artery, and left atrium rise and fall in parallel as the blood volume is changed by bleeding or transfusion (Fig 3 28)

While this leads to the conclusion that the venous system operates in a passive manner, it must also be emphasized that this may be true only if the induced changes in central blood volume are moderate or gradual and not in the nature of emergency reactions It is easy to demonstrate that a fixed pressure-volume relationship need not apply to the veins and venules for they are the most sensitive of all vessels to all types of mechanical, chemical, and temperature impacts For example, the trauma attendant to dissection or to needle insertion may completely close a medium-sized vein In addition, they are supplied with vasomotor nerves through the action of which the distensibility and capacity of the venous system can be actively and greatly changed This provides a means by which the volume of the venous system can change within wide limits without significant corresponding changes in venous

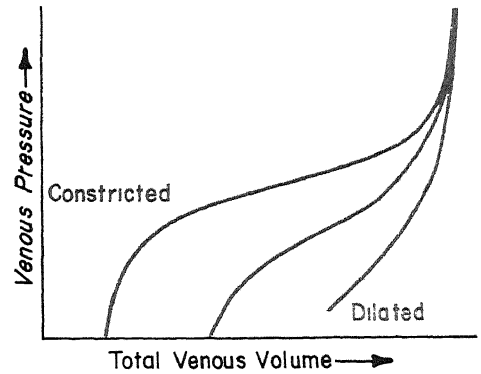


Fig 3 27 Venous distensibility patterns with different degrees of vasomotor tone (Alexander 1954b)

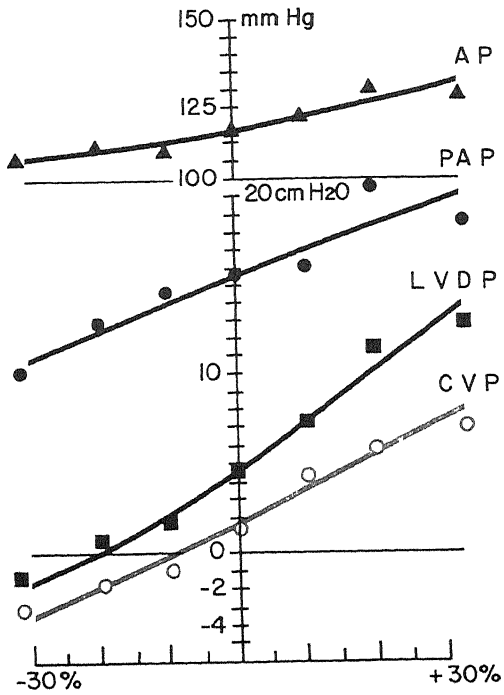


Fig 3 28 Effects of changing total blood volume (by transfusion (+) or bleeding ( )) upon aortic (A P) pulmonary artery (P A P) left ventricular diastolic (L V D P) and central venous pressures (C V P) (Gauer and associates 1956)

pressure Such innervation has been demonstrated for the vena cava and veins of the hind limb and mesentery in animals Active contraction or relaxation of the venous smooth muscle (venoconstriction or dilation) can be elicited by activation of local stretch receptors in the venous walls, and reflexly by activation of the baroreceptors (stretch receptors) in other veins and in the aorta and carotid arteries, by hypoxic stimulation of the carotid chemoreceptors, and by direct local action of adrenergic or cholinergic substances (Fig 3 27) The reflex venoconstriction produced by carotid artery occlusion in dogs can be abolished by the administration or reserpine or guanethidine, suggesting that these reflexes are mediated through the sympathetic nervous system Reflex venomotor activity has also been measured in normal subjects by means of pressures developed within isolated segments of superficial veins An arm superficial vein segment without tributaries is temporarily isolated by applying pressure at the segment ends so that the volume is held constant and vein contraction can be reflected by an increase in pressure which is recorded continuously by an indwelling needle connected to a pressure manometer Changing from the supine to the erect posture results in venoconstriction, as indicated by an increase of the venous pressure within the isolated segment, on resuming the supine position venodilatation occurs Reduction of the circulating blood volume by means of phlebotomy also produces an increase of the venous tone Strong contractions are caused by the Valsalva maneuver, cold, exercise, and hyperventilation Persons with postural hypotension or with absent circulatory reflexes (as in tabes dorsalis) show little or no segment or central venous response to these stimuli, indicating impaired reflex venomotor activity Although the entire venous bed may be regarded as a large variable blood depot, the central systemic veins constitute the most important part of this depot for supplying the right heart It is estimated from observations in dogs that the pressure-volume curve of the venae cavae in the intact living dog can be varied by 15 to 20 times and with only a small effect on venous pressure This means that translocation of large volumes of blood to the pulmonary and systemic arterial circuit is easily effected

VELOCITY CHANGES

In Different Vessels (Fig 3 2B) Just as the cross-sectional area varies greatly in different parts of the cardiovascular system, so does the velocity of blood flow Since the velocity of flow is inversely proportional to the cross-sectional area of the vessel through which it flows, velocity will be maximal in regions of the circulation with the smallest total cross-sectional area Hence, the blood velocity in the aorta is the most rapid of that in any part of the body, but even here the average velocity of about 40 cm /sec , is low In the capillaries, it is least, at about 0 07 cm/sec This slow rate of passage in the peripheral capillaries, of course, provides time for the exchange of materials across the capillary membranes or walls After passing into the veins, the blood again accelerates as the cross-sectional area progressively decreases However, since the caliber of the veins is larger than that of the corresponding arteries, the velocity of venous blood flow never becomes equal to that of the arterial blood and generally equals about one-fourth of that in the arteries For example, the flow of blood along the two venae cavae which together have a cross-sec-

tional area some four times as great as that of the aorta, is only about 8 to 10 cm /sec under resting conditions

Although the average blood velocity is low the peak velocity of blood flow may be high since this changes greatly from moment to moment during each heart beat in the systemic arterial system In the ascending aorta near the aortic valves, the peak velocity during systole may be 120 cm /sec , during most of diastole, the velocity may approximate zero Further down the thoracic aorta and in the abdominal aorta, the small backflow present in the ascending aorta at time of closure of aortic valves is greatly accentuated and extended, while forward flow during diastole is increased In the major aortic branches these violent fluctuations of velocity, although somehwat damped, are largely retained However, except for the coronary, femoral, and axillary arteries, backflow disappears, and the flow curve is often patterned after its corresponding pressure pulse At the capillary level, these fluctuations in velocity are pretty well removed so that a smooth rate of flow is attained In the venules and as the veins become larger, fluctuations or variations in the velocity of flow again reappear In the venae cavae near the right atrium, blood velocity may become much greater during ventricular systole than during its diastole, indicating that contraction of the myocardium itself draws blood toward the heart, not to be confused with *vis a fronte* or cardiac suction (see p 3-103) During inspiration, the whole flow curve in the venae cavae is greatly elevated

Across a Blood Vessel Finally, the velocity of flow will vary in different portions of the cross-sectional area of a blood vessel, especially in the larger vessels (Fig 3 29) In the larger vessels there is a cylindrical layer of blood which adheres to the vessel wall and wets it Inside this motionless layer is another cylindrical fluid layer which flows and rubs against it, and inside this there is another flowing faster, and so on, until the central axial core of blood is reached which has the maximum velocity In arterioles and vessels of quite narrow caliber, red cells are concentrated pretty much in the axial part of the blood stream so that plasma flows along the walls As a result, the red cells move faster than plasma by a significant amount, and hence, the hematocrit in capillaries is somewhat less than that of the larger vessels The effect of this is, of course, to change the distribution of the velocity in the various parts of a cross-section of such a vessel

The flow of blood is laminar or streamlined in most portions of the circulatory tree When the velocity of flow is increased above some critical level turbulence develops

Reynolds' or Random Turbulence The factors determining whether flow is laminar or turbulent include the average velocity of flow, $\bar{v}$ (cm /sec) the vessel diameter, D (cm), the density of the fluid, ρ (about 1 05 g /cm ^{3} for blood) and the viscosity of the fluid, η (about 0 035 to 0 05 poise for blood) The conditions required for laminar flow to become turbulent can be defined when a dimensionless quantity known as *Reynolds' number*, N_R, exceeds a critical value in accordance with the following relation

$$\text{Revnolds' number} = N_R = \bar{v}\, D\rho/\eta$$

The critical value of Reynolds' number for blood flowing through rigid tubes is approximately 2000 In man this value is not exceeded in most parts of the circulation (see Table 3 4), except in the aorta and pulmonary artery where values of 3600 to 5800 or higher may be reached (Schultz et al , 1969) High rates of flow (as in exercise) or decreases in viscosity (e g , in anemia) favor turbulent flow, since cardiac murmurs may be heard over these large vessels during systole in high flow states or anemia, they have been attributed to turbulence However, it is doubted that random turbulent flow can produce sufficient acoustic energy to be audible at the thoracic surface

Aeolian Tones or Fluttering Turbulence and Vortex Formation The forces which produced the musical tones of ancient wind harps of the Greeks probably account for murmur production (Bruns, 1959) As fluid flows past an obstruction or through a stenotic area, vortices or

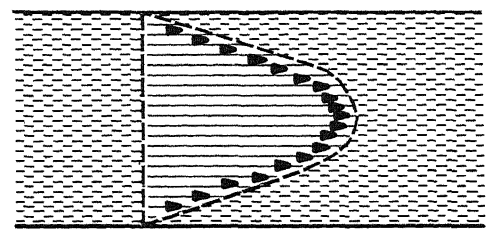


Fig 3 29 Schematic representation of the variable velocity of blood flow in vessels Flow is fastest in the central axial stream and progressively decreases to zero at the vessel wall

swirls are formed which pass downstream as a fluctuating wake Such periodic wake fluctuations generate vibrations in the audible frequency range This kind of turbulence, rather than the random turbulence of Reynolds', probably accounts for most cardiovascular murmurs

PRESSURE CHANGES

Kinetic and Potential Energy Lateral and End Pressure Of the energy developed by the left ventricle, more than 98% is normally in the form of potential energy, 1% or less as kinetic energy When the left ventricle pumps out blood, it increases the radius and length of the aorta and its large branches to store potential energy as tension in the arterial walls Because of the motion (velocity) of the blood during ejection, a certain amount of kinetic energy is also transferred to the blood During exercise with a very large cardiac output or in the presence of aortic stenosis, as much as 10 to 50% of the energy expended by the heart is used for increasing blood velocity Ultimately, however, most of the potential energy is changed to kinetic energy to produce flow through vascular elements, and then is finally dissipated as heat through friction The kinetic energy of flow can have a real effect on intravascular pressure This can be demonstrated by inserting into the flowing bloodstream small catheters or cannulas connected to pressure manometers (Fig 3 30) With the catheter opening pointing upstream, end pressure or potential energy is measured, with the opening at right angles to direction of flow, *lateral* pressure or the pressure actually pressing on the vessel is measured In the vessels of humans, end pressure and lateral pressure are almost equal except in the aorta and its immediate branches in which the kinetic energy is sufficiently high from prevailing blood velocity to cause the lateral pressure to be significantly lower, i e , lateral pressure is inversely proportional to the blood velocity (Bernoulli effect) At rest, the average aortic lateral pressure is about 1 mm less than the end pressure, with exercise, the difference might be 15 mm Hg If one is recording phasic aortic pressure throughout the cardiac cycle, then with exercise the lateral pressure might be 40 mm Hg less than the end pressure during the period of rapid blood ejection (first third of systole) Ultimately, most of the potential energy is converted into kinetic energy in producing flow In this exchange no energy is lost However, as

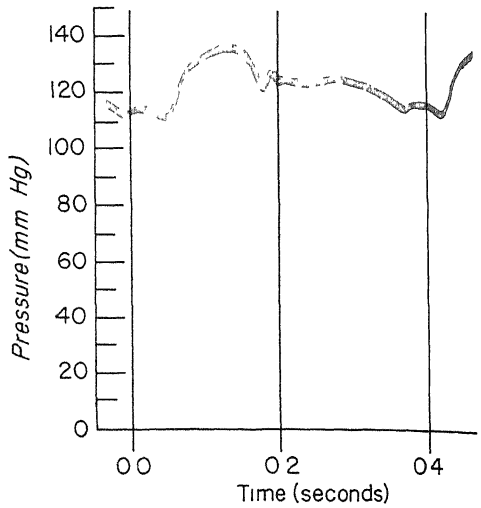


Fig 3 30 Reproduction from original record of lateral upstream and downstream phasic pressure patterns taken at the subclavian level in the aorta of a large open chest dog by means of three strain gauges adjusted to identical zeros and sensitivities The patterns and pressure values of the three curves are almost identical at systolic pressure 139 mm Hg diastolic pressure 115 mm Hg heart rate 150 min

blood progresses from the aorta to the atrium, the intravascular pressure decreases to very low values This permanent loss of energy or pressure arises from fluid friction developed within the vessels with gradual change of energy into heat

In different vessels (Fig 3 2C) Pressure in the aortic arch reaches a maximum during mid systole, and a minimum at the end of diastole The numerical difference between these pressures is called the "pulse pressure " Similarly, pressures in the thoracic and abdominal aorta and their major branches, i e , the carotid, the radial, and especially the femoral, show that the values for the systolic pressure increase and those for diastolic pressure decrease, i e , pulse pressure increases In still smaller arteries, the pulse pressure progressively diminishes, systolic pressure falling more than diastolic However, this does not represent a frictional loss of energy but merely a replacement or shift in the temporal relations of pressure during a cardiac cycle, since the mean pressure, for example, in the femoral arteries is close to that in the aorta

As the arteries divide and subdivide, the caliber of the individual vessels diminishes and the pressure gradients become correspondingly steeper, particularly in the arterioles and capillaries Not only is the pressure gradient steeper,

but the oscillations of pressure during a cardiac cycle are reduced to a minimum at the time the arterioles are reached At the arterioles there is a pronounced decrease in pressure gradient, a second abrupt decrease in the capillaries with almost no residual pressure fluctuations A more gradual pressure decline occurs in the venules and veins until a negative pressure is actually reached in the central veins owing to the effect of the subatmospheric pressure within the thorax

Functional Characteristics of the Pulmonary Circulation

In many respects the pulmonary vascular bed (pulmonary artery through pulmonary veins) is similar to the systemic circuit but certain differences deserve emphasis The pulmonary circulation is a low resistance circuit, arterioles being essentially absent, the pulmonary vessels supply only one type of tissue (alveoli) so vasomotor requirements are low, pulmonary blood volume is less than that in the systemic circuit, extravascular conditions (within the thoracic cage) are fairly uniform (Fig 3 31)

The main pulmonary artery and its branches are structurally similar and of about the same or slightly larger diameter than the aorta, but their walls are one-third the thickness and their branches are much shorter and of very small capacity Small arteries in both systems have thick muscular walls Arteries less than 0 1 mm in diameter are endothelial tubes without muscle There are no vessels corresponding to arterioles, i e , no small bore high resistance vessels The capillaries are somewhat larger, appear to take off very densely from the above endothelial tubes, and have multiple anastomoses There are numerous AV shunts The pulmonary veins are short and contain less blood than the systemic veins

The pulmonary artery and pulmonary bed in general are many times more distensible than the systemic portion of the arterial circuit and probably represent the most distensible part of the low pressure system Simultaneous pressure and diameter measurements of the main pulmonary artery in living, thoracotomized dogs show a linear relationship over relatively wide ranges of pressure up to 50 cm water The pressure and diameter pulse contours are almost identical, indicating minimal inertance and viscous resistance of the vessel wall In vivo pressure-volume curves of the entire pulmonary vascular bed and of the pulmonary veins (and left atrium) have pretty much the shape of venous pressure volume curves in the relaxed state (Fig 3 31) The blood volume of the pulmonary system is only about 10% of the total blood volume of which about 60 to 80 ml is in the capillaries, and the rest is in the major arteries and veins Since the pulmonary vessels are a low pressure distensible system, a minute increase in outflow pressure at the left ventricle or an increase in pressure at the right ventricle results in a sizeable accumulation of blood in the lungs Thus, the blood volume of the pulmonary vascular bed has an important reservoir function and is about the most important depot for supplying the left heart

The pulmonary circulation has been considered as a variable blood reservoir which serves to supply the left ventricle with a steady supply of blood, thus buffering changes in venous return to the right heart It was believed to function more or less passively to shifts in blood volume caused by systemic venoconstriction, changes in blood volume and variations in ventilation This is an oversimplication (Heinemann and Fishman, 1969) The pulmonary circulation is under sympathetic control and as sympathetic tone increases, the compliance of

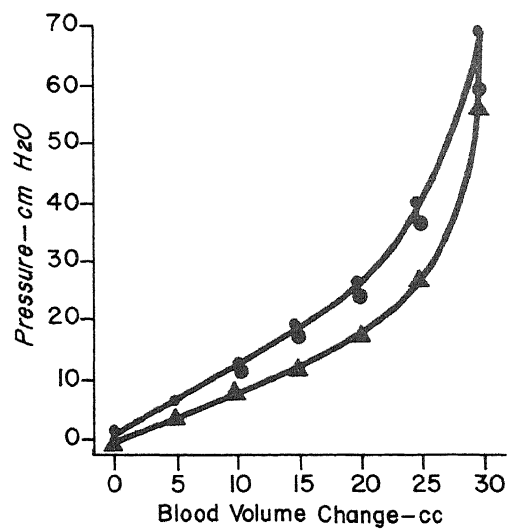


Fig 3 31 Pressure volume relationship of the pulmonary vascular bed Injections and withdrawals made at 10-sec intervals Dots pressures immediately after end of each injection Closed circles pressures 5 sec after end of each injection Triangles pressures immediately after each withdrawal From Sarnoff and Berglund 1952)

the pulmonary vessels decreases This improves pulsatile blood flow transmission from right to left ventricle (see p 3-125) Changes in tone of the pulmonary vascular tree, thus, play a role in maintaining appropriate delivery of blood to the left ventricle and are probably more important than reservoir function in rapid circulatory regulations In pathological conditions, the blood reservoir function is obvious Thus in left heart failure pulmonary blood volume increases, while it diminishes in hypovolemic shock

Since the stroke volume and mean velocity of flow are almost identical in the aorta and pulmonary artery, the kinetic energies of the two bloods are the same, but because the potential energy imparted to the blood by the right ventricle is much less, the kinetic factor is a greater proportion of the energy output of the right ventricle Phasic variations in blood flow velocity are possibly even more marked in the pulmonary artery than in the aorta In contrast to the systemic circulation, a markedly pulsatile flow is retained in the pulmonary capillary bed during rest and after exercise During systole, blood is accelerated in the right ventricle and pulmonary artery by contraction of the right ventricle The blood gains kinetic energy or energy of motion As blood distends the arterial tree, the latter acquires potential energy (elastic and hydrostatic) At the same time, energy is partially dissipated by resistance to systolic flow in the small vessels During diastole the potential energy in the arterial tree is partly lost in hysteresis, and the remainder is dissipated in elastic recoil as blood flows through the resistance of the small vessels By the end of diastole, the pressures in the vascular bed are almost equilibrated and capillary flow practically ceases

The resistance to blood flow in the pulmonary circulation is about one-thirteenth that of the systemic circulation This is because there are no high resistance arterioles, the pulmonary capillaries are large and very numerous, the pulmonary vessels are so easily distended passively, and large portions of the pulmonary bed are only partially used normally The pressure pulsation in the pulmonary artery of 22/8 (mean 13 mm Hg) is fairly well maintained in the capillary bed and pulmonary veins since the capillary flow is so pulsatile during a cardiac cycle However, the capillary pressure has not been directly measured Phasic pressures taken indirectly through a catheter passed through the pulmonary artery until it is "wedged" in its small branches have been assumed to represent "capillary" or left atrial pressures since the blood obtained through the catheter is well oxygenated However, what such recorded phasic pressure curves represent has never been clearly defined The pressure gradient through the pulmonary bed is quite small Its highest pressure (pulmonary artery 13 mm Hg mean) drops to about 7 mm Hg in the left ventricle at the end of diastole and at the site of outflow from the pulmonary circuit This pressure gradient of about 6 to 7 mm Hg causes to flow through the pulmonary circuit the same volume of blood as flows through the systemic system with a pressure gradient of about 90 mm Hg

The capillary pressure in the pulmonary circulation is far lower than in most parts of the systemic circulation (e g , 7 mm Hg) while the plasma colloidal osmotic pressure is at its usual value (e g , 30 mm Hg) This has led to the postulate, supported by some experimental evidence, that a large negative interstitial space pressure (e g , −15 mm Hg) must prevail to account for pulmonary capillary exchanges (Guyton, 1969)

In addition to its function in blood gas exchange and as a blood volume reservoir, the pulmonary circulation serves to filter the systemic venous blood Small foreign particles, blood clots and other types of emboli are trapped in the lungs when they are removed by proteolytic enzymes and phagocytosis Since the pulmonary arterial system and bronchial arterial system share a common capillary network, embolization of a small pulmonary artery does not ordinarily abolish the blood supply through alveolar membranes Thus, the affected lung tissue survives until the embolus is removed When large emboli occlude major arteries, the bronchial circulation is inadequate to prevent necrosis Since most emboli enter the systemic venous system, the filtering function of the pulmonary circulation protects such vital organs as the brain, heart, and kidneys from embolic infarction White blood cells and platelets are also removed from the blood in the pulmonary circulation as are erythrocytes after incompatible blood transfusions

Interrelation between Pressure, Flow and Resistance

Hemodynamics is the study of the forces which cause, alter, and regulate the flow of blood through the body

A simple association of pressure and flow can obviously serve as an index of total peripheral resistance (TPR) to flow in the cardiovascular system, thus

$$\frac{\text{(Aortic Pressure}}{\text{(Aortic flow}} \text{ or } \frac{\text{Pulmonary artery pressure)}}{\text{Pulmonary artery flow)}}$$

or to local resistance in single bed or organ, for example

$$\frac{\text{(Renal artery pressure)}}{\text{(Renal artery flow)}}$$

Such measurements and calculations have been used extensively in recent years in clinical and experimental investigation and, especially in the aorta, to separate the resistance changes due to central or cardiac factors from those due to peripheral factors The architecture of the whole circulatory distributing system is such that the organs are supplied through many parallel shunts In this the total peripheral resistance (TPR) is

$$(1/\text{TPR}) = (1/r1) + (1/r2) + (1/r3) \quad (1/r^n)$$

but since the aorta supplies all the regions with pressure and flow, TPR can be calculated as mean aortic pressure divided by cardiac output (see p 3–192) A little reflection will indicate, however, that the resistance to flow can arise from active changes in the smooth muscle of the particular vascular bed in question, or it could be related to passive changes in the geometry of the same vascular bed due to internal and external variations of the pressure The above approach does not permit separation and quantitation of active and passive changes in vessel size which is essential to the proper understanding of pressure and flow relations In a modified form, some physical laws derived from hydrodynamics have application to the flow of blood, and they will be considered next

POISEUILLE'S LAW

Generally stated, the volume (F) of blood flowing through a circulating system increases with the perfusing pressure (P) and decreases with the resistance (R) to flow as indicated by the equation $F = P/R$ This is Poiseuille's experimental law which is expressed in more detail as

$$F = (P_1 - P_2)\ \pi r^4/8L\upsilon$$

in which F is flow per unit of time, P_1 is pressure at an upstream point in the tube, P_2 is the pressure at a downstream point in the tube, π is 3 1416, r is the tube radius, L is tube length, υ is fluid viscosity Poiseuille's experiments were based upon measurements of streamlined flow (nonturbulent) in nondistensible tubes with Newtonian fluids at constant temperature This means that flow is augmented by an increase in any factor in the numerator and decreased by any factor in the denominator Thus, flow would decrease by one-half if tube length is doubled or if blood became twice as viscous Blood is caused to flow from one point to another in a tube by the difference of pressure between such points This means that doubling the pressure gradient ($P_1 - P_2$) along the tube will double the flow Change of radius is much more effective in changing flow, for doubling the radius will increase flow 16 times

The resistance to flow, defined as the ratio of the driving force to the flow, is obtained by a rearrangement of the above formula in which

$$R = (P_1 - P_2)/F = 8L\upsilon/\pi r^4$$

Resistance to flow is made up of two distinct factors These are the viscosity (υ) or inner friction of the fluid and the geometry of the tube

$$(8L)/(\pi r^4)$$

which is made up of the radius and length Thus the terms on the right of the foregoing equation represent the frictional load affecting liquid flow through a straight, rigid, narrow tube In his system, the resistance was constant since both the viscosity and the geometry of the tube were unchanged and did not change with rate of flow or pressure A plot of flow versus pressure gave a straight line through the origin, and one of resistance vs pressure a horizontal straight line (Fig 3 32)

APPLICATION OF POISEUILLE'S LAW TO THE CIRCULATION

The applicability of this law in toto to the circulation depends upon whether resistance is independent of pressure and flow In the circulatory system peripheral resistance is a measure of the totality of all factors affecting blood flow. These include change (decrease) in apparent blood viscosity (which is known to occur with increase in perfusion pressure, occasioned by the movement of the red cells to form a central axial rod) the existence of streamlined vs turbulent flow, the length of vessels and the cross-sectional areas of the blood vessels The latter

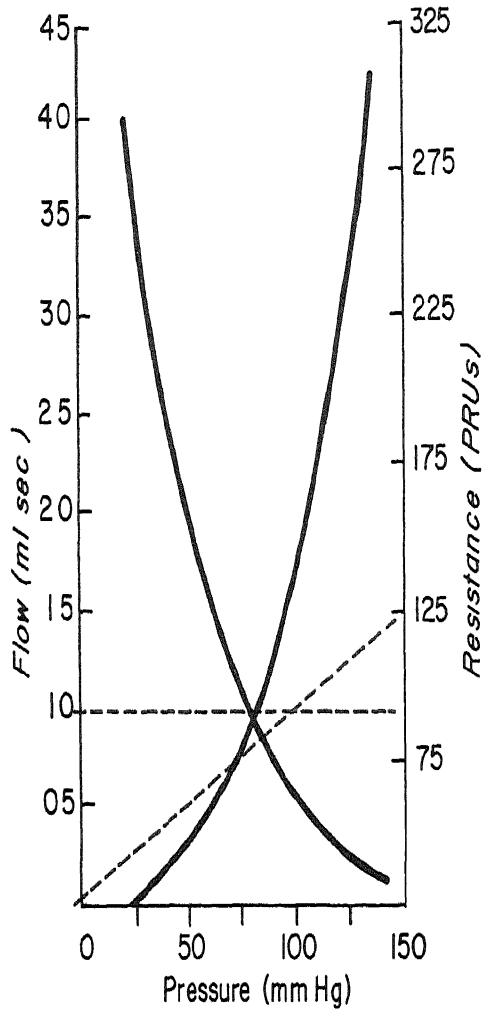


Fig 3 32 The relationships of flow pressure and resistance in nondistensible tubes using Newtonian fluids (*dotted lines*) at constant temperature (Poiseuille s law) and blood in vessels (*solid lines*) See text for details

is determined by the extravascular pressure provided by surrounding tissue, by mechanical dilation with rising perfusion pressure, by opening of new capillaries and vessels with change in metabolism and with rising perfusion pressure, and by active changes in the state of contraction of the muscular walls through vasomotor nerves, humoral substances and metabolic products

The flow-pressure curves obtained with blood in vascular beds (Fig 3 32) are quite different from those of the artificial system used in Poiseuille's experiments The chief differences are the curves do not pass through the origin, and flow is insignificant until a critical level of pressure is attained (called by Burton the critical closing pressure) The curves for dilated vessels are almost linear, but with increasing tone the curves become sigmoid and then convex to the flow axis At sufficiently high pressures, the curves tend to become straight lines pointing toward the origin The resistance is almost constant at high pressure but rises as the pressure is reduced to reach very high values near the critical closing pressure Such deviations could arise if *1)* the geometric factor of Poiseuille's law is not constant but varies with the distensibility of the blood vessels, and *2)* blood has anomalous viscosity so that the viscosity factor of Poiseuille's equation varies with the rate of flow and, hence, the driving pressure While experimental evaluation is difficult, it does appear that vessel distensibility is the major factor responsible for the deviation of in vivo pressure-flow curves from Poiseuille's law This in turn is related to the active tension or contraction of smooth muscle in the vessel walls and to the combined effects of elastic and collagenous fibers in the architecture of the vessels

2 / THE HEART: GROSS STRUCTURE; MYOCARDIAL CELLS

Weighing some 330 g and about the size of a large man's fist, the heart is a four-chambered double pump Each 24 hr it contracts more than 100,000 times, pumping over 60 barrels (7200 liters) of blood

Gross Structure

Some of the important structural features of the heart are depicted in Figure 3 33*A* The fibrous skeleton (Fig 3 33*B*, *C*) is a dense connective tissue framework separating the muscular layers of the atria from those of the ventricles It forms the annuli fibrosi, or fibrous rings, which surround the orifices of the aorta, pulmonary artery, and the atrioventricular valves These rings provide firm attachment sites for the valves of the heart and for the insertion of the cardiac muscles

THE ATRIA

The atria are relatively thin-walled structures which act as receiving chambers for the venous blood returning to the heart The myocardium is disposed in two sheets A superficial or oblique sheet continues over both atria Internal to this are deeper layers, one for each atrium, in which the fiber bundles are arranged more or less at right angles with the superficial bundles Certain features of the atria which are of physiological significance are depicted in Figure 3 34

Three functions are subserved by the atria (Mitchell and Shapiro, 1969) *1*) elastic reservoir and conduit from the venous bed to the ventricle, *2*) booster pump action, enhancing ventricular filling and maintaining low mean atrial pressure, *3*) atrioventricular valve closure before ventricular systole Following atrial contraction, which increases ventricular pressure, pressure in the relaxing atria declines The resultant ventriculo-atrial pressure gradient aids or actually effects atrioventricular valve closure (Patterson et al , 1965)

The pumping function of the atria is relatively unimportant at slow heart rates The ventricles fill, for the most part during the period when atria and ventricles are quiescent and venous blood flows into the heart as into a common chamber At the very end of this period, the atria contract and increase ventricular filling In normal subjects atrial contraction increases cardiac output from 15 to 20% and from 20 to 30% in patients with certain heart diseases (Benchimol, 1969) (see also p 3-104)

THE VENTRICLES

As can be seen from Table 3 5, the ventricular myocardial mass makes up most of the heart weight Four major bands of muscle may be identified by dissection (Fig 3 35) *1*) the superficial sino spiral (with origin and insertion at the tricuspid valve ring), *2*) the superficial bulbo spiral (with origin and insertion at the mitral valve ring), *3*) deep sino spiral muscles which encircle both the ventricles, *4*) deep bulbo spiral muscle fibers which encircle only the left ventricle

The deep muscle fibers spiral very gradually so that they actually act as circular constrictor muscle fibers Although these muscular bands may be dissection artifacts, the transition of fiber direction from counterclockwise to clockwise across the ventricular wall is clear (Grant, 1965)

When the ventricles contract during the so-called isovolumetric contraction period, the chambers are closed and no reduction in heart volume occurs until the pressure is great enough to open the valves leading to the pulmonary artery and aorta During this period, the heart changes in shape from an ovoid to a more spherical form Then, as blood is ejected into the aorta and pulmonary artery, the transverse diameter decreases and the base of the heart moves downward, shortening the distance from base to apex Since the spiral muscles

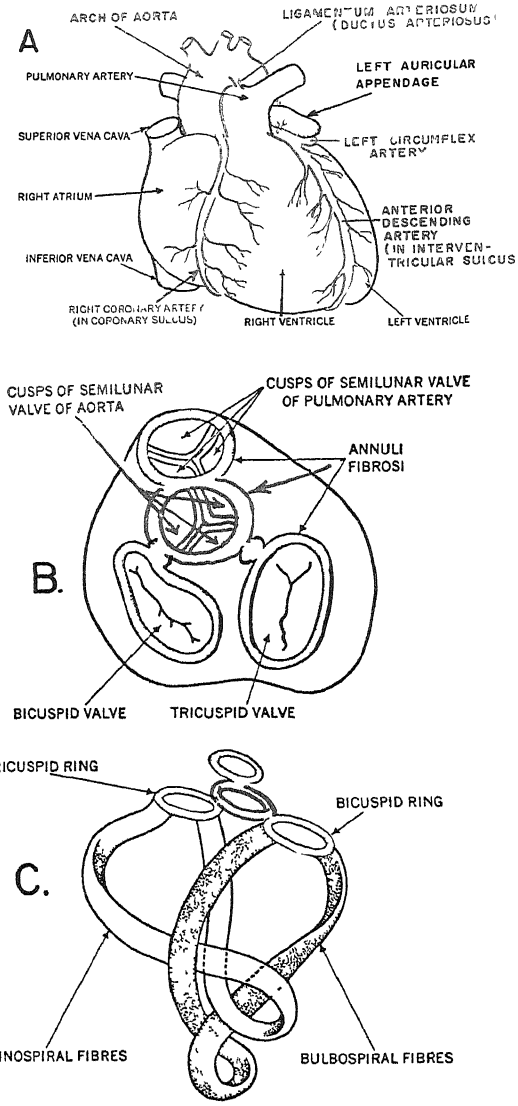


Fig 3 33. ***A*** front view of the human heart ***B*** fibrous skeleton and valves of the heart viewed from above ***C*** spiral muscles of the ventricle (Redrawn from Truex 1967)

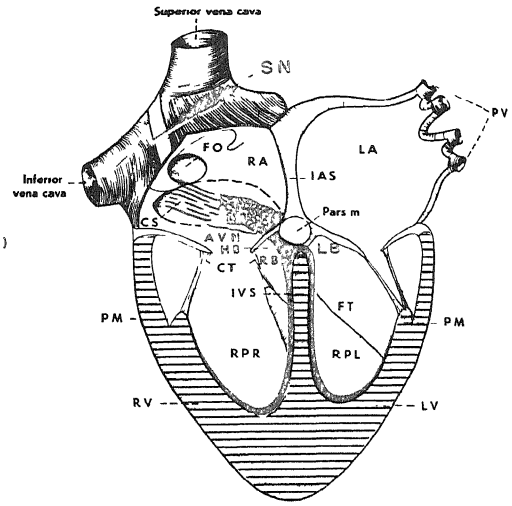


Fig 3 34 Specialized structures and anatomical landmarks within the mammalian heart S N sinoatrial node P V pulmonary veins R A L A right and left atria I A S interatrial septum C S coronary sinus A V N atrioventricular node H B bundle of His R B L B right and left branches of the bundle of His I V S inter ventricular septum P M papillary muscle F T false tendon of the Purkinje network R V L V right and left ventricle (Redrawn from Schutz 1958)

Table 3 5
Relative weights of the heart and various myocardial chamber walls

Body weight	54 7 kg
Heart weight	297 g
Heart weight as percentage of body weight	0 54%
Left ventricle	85 5 g
Right ventricle	49 7 g
Interventricular septum	63 6 g
Left atrium	17 0 g
Right atrium	12 5 g
Interatrial septum	9 5 g

After Bargmann, 1963

turn in opposite directions, there is little rotation of the heart with contraction, save for a slight turning to the right

Thus, the volume of the left ventricle is reduced by a decrease in the transverse diameter caused by shortening of the constrictor fibers and shortening of the base-apex diameter The left ventricle can be thought of as a cylinder which displaces volume by decreasing in height and by reducing its diameter This latter effect is particularly effective owing to the constrictor effect of the deep muscular band

The right ventricle is much thinner walled than the left ventricle (see Table 3 5) Its free wall is moved toward the interventricular septum by contraction of the spiral muscles If the entire free wall of the right ventricle is rendered inactive by cauterization, some reduction in volume will still take place during systole As the constrictor muscles of the left ventricle contract, the curvature of the septum is increased, pulling the free right ventricular wall toward the septum This process is called *left ventricular aid* (See page 3-99)

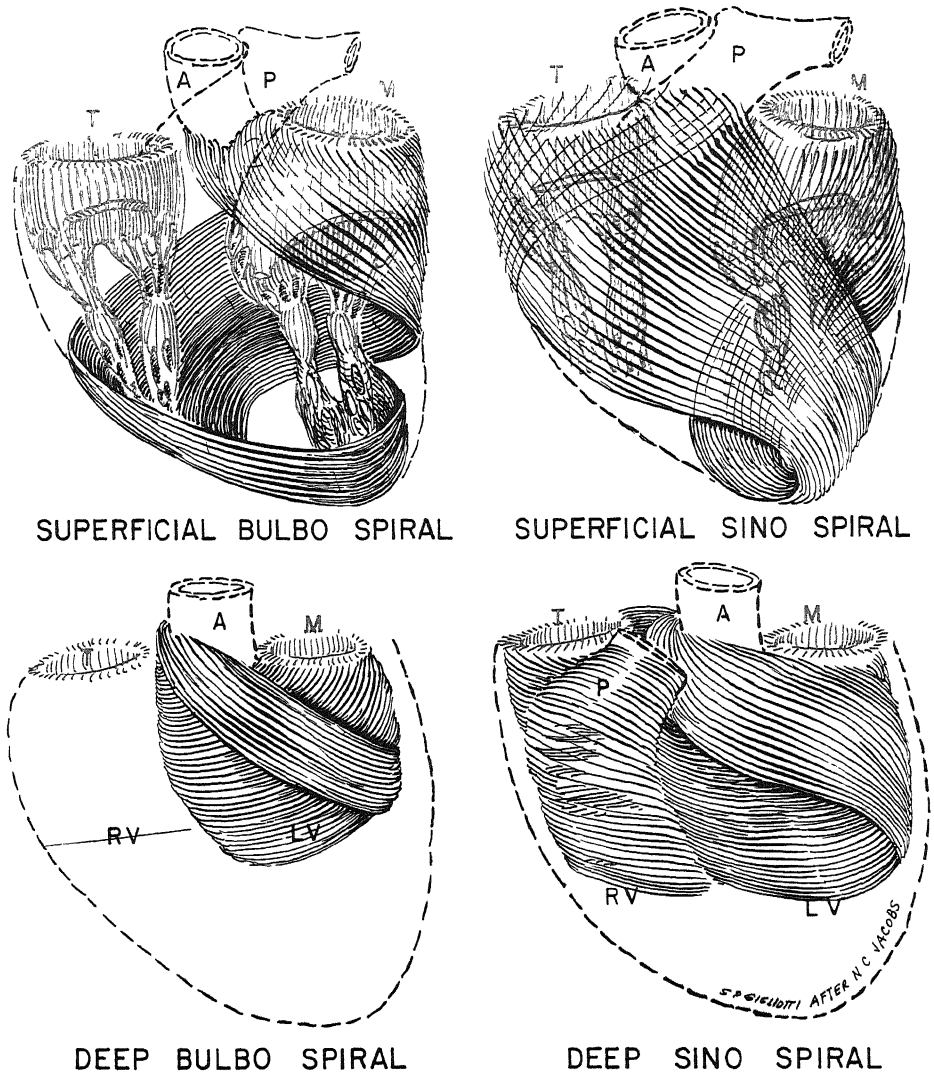


Fig 3 35 Demonstration of the superficial and deep ventricular myocardial bundles (Robb and Robb 1942)

ROLE OF THE PERICARDIUM

The pericardium is a double-walled sac containing a few cubic centimeters of serous fluid which gives a smooth lubricated surface for the heart's movements Although its potential volume is larger than that occupied by the normal heart in diastole, still, after moderate enlargement of the contained heart, the intrapericardial sac pressure rises greatly with small volume influence Thus, in the case of left ventricular stress and enlargement, the pericardium restrains diastolic expansion, especially of the right ventricle, and, therefore, limits right ventricular work This protects the left ventricle from overload and the pulmonary circulation from congestion The pericardium also protects against tricuspid or mitral valve regurgitation at high ventricular filling pressures The value of these mechanisms is probably limited to situations involving acute changes in heart volume, when cardiac enlargement develops more gradually, the pericardium follows suit and stretches to accommodate the larger contents Congenital lack of a pericardium is not uncommon in animals and humans who live to a ripe old age without evidence of cardiac distress

Myocardial Cells

Functionally, there are two general types of myocardial cells cells specialized for impulse formation and conduction, and cells specialized for contraction Cardiac tissue is composed of discrete cells, each bounded by an intact plasma membrane However, these anatomically distinct cells form a functional syncytium because of the presence of "tight" junctions characterized by low electrical resistance (see intercalated discs) These allow the passage of certain ions and perhaps small molecules from cell to cell

The cells specialized for pacemaking and conduction functions are capable of generating spontaneous rhythmic electrical potentials Those specialized for contraction remain at rest until stimulated by a conducted impulse which has arisen elsewhere

The individual myocardial cells in all mammals (mouse through whale) are about the same size (10 to 15 μ in diameter) Thus, the hearts of larger species are composed of many more cells than those of smaller species (Sommer and Johnson, 1970)

PACEMAKER AND CONDUCTING CELLS

Three types of cells specialized for impulse formation and rapid conduction have been described (James and Sherf, 1968) P cells, transitional cells, and Purkinje cells

The *P cells* have been so named because *1)* they are pale, *2)* they are primitive in internal structure compared to the more highly differentiated contractile cells, and *3)* pacemaking appears to be their primary function They are located in the sinoatrial and in the atrioventricular nodes These cells, on cross-section, are oval or rounded, in contrast to the elongated shape of other myocardial cells, and their surface is smooth They measure 5 to 10 μ in greatest diameter and contain relatively few intracellular organelles

Transitional cells are slender and elongated They are both shorter and narrower than other myocardial cells Their internal complexity is intermediate between the P cells and the contractile myocardial cells They appear to form the only connection between P cells of the sinoatrial node and the remainder of the heart

The *Purkinje cells* measure from 10 to 30 μ in cross-section and 20 to 50 μ in length They are shorter and broader than working myocardial cells and more commonly joined end to end, rather than side to side

CONTRACTILE OR WORKING MYOCARDIAL CELLS

These form the bulk of cardiac cells and are highly organized for their contractile function For the most part they are joined longitudinally in series with multiple cells forming the so-called myocardial "fibers" seen under the light microscope These "fibers" are in turn arranged parallel to one another In contrast to skeletal muscle, they have many lateral and end to side connections joining the separate plasma membranes of each cell Unlike the sparse linear myofibrils of Purkinje cells and the sparse and randomly distributed myofibrils of P cells, the myofibrils of working myocardial cells are densely packed Individual cells are some 10 to 20 μ in diameter and 50 to 100 μ in length

Working myocardial cells are architecturally similar to skeletal muscle cells but exhibit a number of important functional and structural differences The major morphologic features are diagrammed in Figure 3 36

Fiber, Fibril, and Filament. The myocardial cell is a specialized striated muscle cell containing a centrally located nucleus, a continuous limiting plasma membrane (the sarcolemma), and numerous contractile myofibrils which are separated by varying amounts of sarcoplasm A mucopolysaccharide-rich basement membrane invests the sarcolemma of each cell The intercalated discs join individual cells end to end The term *fiber* is sometimes used to designate an individual myocardial cell but is more commonly applied to a series of cells joined end to end which thus form a long fiber The fibers are separated from one another by intercellular spaces containing collagen, fibroblasts, and capillaries Individual cells frequently branch to interdigitate with adjacent fibers, forming a continuous network of cells Groups of muscle fibers are invested by sheaths of connective tissue separating them into long bundles Individual bundles also interconnect with other bundles The muscle fibers within a given bundle are parallel, but adjacent bundles may be oblique or transverse to one another The cells contain longitudinally arranged myofibrils which extend their full length and insert into the cytoplasmic surface of the intercalated discs These multiple parallel myofibrils are cross-banded and each is divided into a series of repeating units, the sarcomeres, the fundamental structural and functional units of

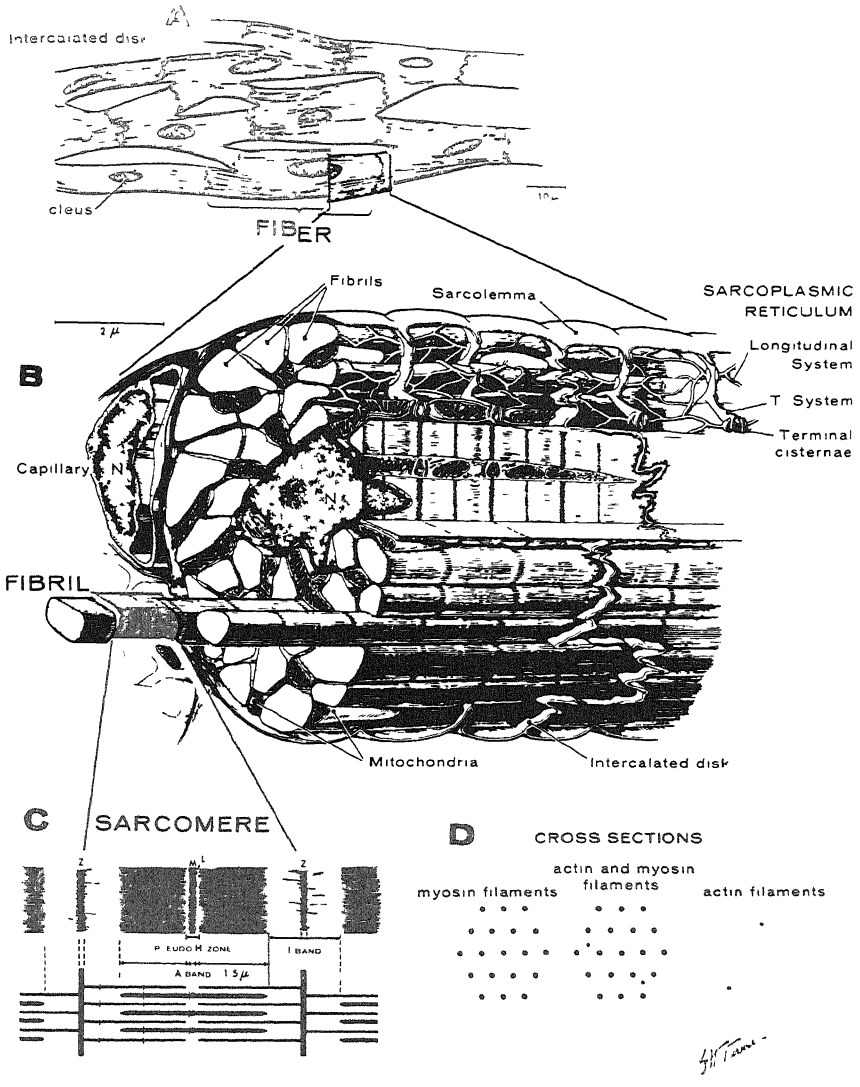


Fig 3 36 Structure of the myocardium as seen with light (*upper*) and electron (*lower*) microscopy In the diagram of a sarcomere the various structural features seen on longitudinal section are identified The sarcomere is delimited by two dark Z lines The I band is a region on either side of the Z line containing only thin actin filaments These extend into the lateral portions of the A band where they interdigitate with the thicker myosin filaments At the center of the A band is a dark line the M line flanked by two lighter areas the L lines This M-L complex is called the pseudo-H zone When the actin filaments do not penetrate as far as the M line the band where there is no overlapping of the actin and myosin filaments is designated the H zone This zone is seen only when the sarcomere length exceeds 2 2 μ (From Braunwald et al 1967)

contraction These sarcomeres make up about one half of the cell mass and are aligned in transverse register, i e , the ends of the sarcomeres in adjacent fibrils are next to one another and are of the same length It is this arrangement which gives the fiber its banded or striated appearance under the light microscope The sarcomeres, in turn, are composed of filaments (myofilaments) consisting of strands of contractile proteins

The individual sarcomere is composed of a number of proteins, some as yet incompletely characterized The two primary and best defined components of the sarcomere are actin and myosin The sarcomere is delimited at either end by an amorphous, dark transverse

septum called the Z line The area between two Z lines is composed of interdigitated thick and thin myofilaments which overlap in such a way as to produce the alternating light and dark transverse bands The distance between two Z lines varies between 1 5 and 2 2 μ, depending on the overall muscle length At the center of the sarcomere, a broad dark band of constant width (1 5 μ), called the A band, is found This is formed by parallel ranks of thick (100 Å) myosin filaments which are located in the central portion of each sarcomere On either side of the A band are two lighter bands, termed the I bands, which are of variable width These are formed by longitudinally oriented parallel ranks of thin (50 Å) actin filaments which extend from each Z line toward the center of the sarcomere Where they meet the myosin filaments of the dense A band, the two types of filaments are interdigitated for a variable length depending upon the state of contraction In the midportion of the A band, where there is no overlapping of thin and thick filaments, a lighter H zone occurs The central portion of the A band is bisected by a dark line, the M line, which is flanked by two lighter areas, the L lines This central area of the A band is called the M-L complex or pseudo-H zone Thus, the sarcomere consists of an orderly array of partially overlapping rodlike myofilaments These myofilaments are of fixed length but the sarcomere may change its length during contraction and relaxation through an interaction of the actin and myosin filaments These slide past one another producing a greater or lesser degree of overlapping

The myofilaments are made up of aggregates of contractile proteins Actin is composed of two linear chains of globular sub-units wound around the other in a helical arrangement (see Fig 3 37) The myosin units are held together by a lateral arrangement of cross linkages which make up the centrally located M band Each myosin subunit is rod shaped with a globular enlargement at one end They are lined up in the myosin filament with the globular portions clustered at either end and the rods occupying the central part of the filament (Fig 3.37) The globular ends of the myosin molecules consist of the more rapidly sedimenting heavy meromyosin which contains both the ATPase and actin-binding sites of myosin These rounded subunits of heavy meromyosin are thus arranged at either end of the thick

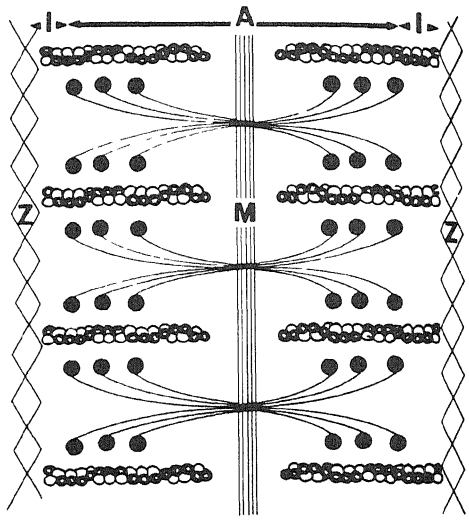


Fig 3 37 Diagram representing the arrangement of actin and myosin in the sarcomere The Z bands probably are made up of a lattice like arrangement of tropomyosin The subunits of actin are globular and linked in two strands which form a double helix In cross-sections at each pitch of this helix is seen the globular end of a myosin filament consisting of heavy meromyosin The subunits comprising each myosin filament are lined up with the globular portions clustered at either end and the rods occupying the central part of the filament Cross bridge formation during contraction probably occurs at these sites The M band at the center of each sarcomere is a system of cross links which fix the myosin units in position (Redrawn from Legato 1969)

myosin filament such that one projects as a cross-bridge just under each pitch of the actin helix Cross-bridge formation occurs at these active sites of the two filaments, and sequential making and breaking of these interconnections produces the sliding of actin along myosin which occurs in muscular contraction When heavy meromyosin is applied to isolated actin filaments they become aligned in a reproducible configuration indicating that both actin and myosin filaments have polarity In sarcomeres it is thought that the actin filaments are so arranged that the polarity of the filaments is reversed on either side of the Z band This orientation would give direction for cross-bridge formation between the myofilaments so that the actin strands at either end of the sarcomere can only move toward one another This arrangement permits shortening of the sarcomere when sequential making and breaking of cross linkages occurs during contraction

Since it has been observed that both the thick and thin contractile filaments of the sar-

comere are constant in length at rest and during contraction, these interdigitating filaments must slide over one another when shortening and lengthening occurs During contraction the actin filaments travel farther into the A band The A band remains constant in length while the I bands become shorter as the Z lines move toward one another

In vivo, the mechanisms involved in excitation and contraction require the presence of at least two other proteins called modulatory proteins (tropomyosin and troponin) The functions of these proteins will be discussed later under the heading "Contractile Proteins"

Intercalated Discs. Adjacent cells in a series are joined by specialized, paired membrane junctions, the intercalated discs The transverse portions of these discs are at right angles to the muscle fibers They are located at the ends of cells and always at the level of a Z line These cell boundaries follow a zigzag path neighboring transverse segments are often connected by a longitudinal segment which runs the length of a sarcomere to the next Z line, thus forming steps between certain cells The intercalated discs constitute structural cell boundaries so that the myocardium is not a true anatomical syncytium However, since there is low electrical impedance between the cells (Weidmann, 1966) and the attachments at the intercalated discs are firm, the myocardium functions as a syncytium both mechanically and electrically

Cell Junctions. Three types of junctional specializations form the intercalated disc (McNutt, 1970) *1) The Macula Adherens or Desmosomes* These are dense, paired sarcoplasmic bodies found opposed on either side of the intercalated disc, not in relation to myofibrils They seem to form an area of strong adhesion between the adjacent cells *2) The Fascia Adherens* This structure is generally larger than the macula adherens and occupies the major part of the transverse segments of the intercalated disc The thin actin filaments of the I band insert into the fascia adherens The thin myofilaments do not appear to insert into the maculae adherentes The fascia adherens appears to provide adhesion between adjacent cells and a locus for insertion of myofilaments on either side of the intercalated disc Thus, together with the maculae adherentes, these structures provide for mechanical attachment of adjacent cells, allowing the transmission of contractile force from cell to cell *3) Nexus* Nexuses or "tight junctions" are generally found in the lateral portions of the disc No filaments insert into them and they seem to be out of the direct line of tension exerted by myofilaments The plasma membrane of one cell comes into contact with that of the adjoining cell at the nexus Nexuses appear to serve as efficient low-resistance pathways for electrical conduction between the cells

Mitochondria. The mitochondria or sarcosomes are remarkably numerous, making up from 25 to 30% of the myocardium They are the major sites of oxidative phosphorylation, in which energy provided from substrate oxidation is converted to adenosine triphosphate, the source of energy for cell function

Sarcoplasmic Reticulum. The sarcoplasmic reticulum of cardiac muscle differs from that of skeletal muscle in several ways The two systems are compared diagrammatically in Figure 3 38 The first element, termed a *transverse component* or *T system*, consists of relatively thick-walled transverse tubules formed by invaginations of the sarcolemma which extend into the myocardial fiber Invaginations of the sarcolemma, continuous with the T system, are seen located at the level of the Z line Although the T tubules are usually found at the level of the Z lines, T tubules may be disoriented to run longitudinally in the fibers (Tice, 1967) Thus, the membrane forming the T tubule is continuous with the sarcolemma and the lumen of the tubule is in direct connection with the extracellular space The second component is termed the *longitudinal system* or *L system* It consists of a series of anastomosing thin-walled tubules which invest the sarcomere These tubules form a complex reticulum of anastomosing intracellular channels having no direct contact with the cell exterior It will be noted in Figure 3 38 that the T tubes of cardiac cells are considerably larger than those found in skeletal muscle and their relationship to the longitudinal tubular system is dissimilar The relationship between the T and L systems in cardiac and skeletal muscle is depicted in Figure 3 39 The cardiac sarcotubules form a less constant and predictable pattern than those of skeletal muscle The latter form repeating units which are contained between two Z bands of a single sarcomere and have a fenestrated cistern at the central portion in the region of the M band (Fig 3 38*B*) Car-

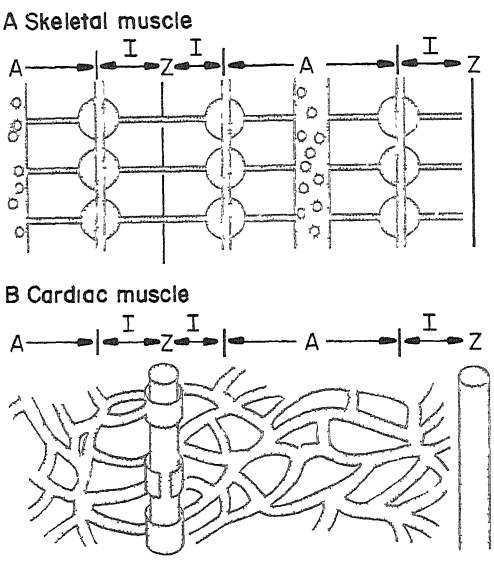


Fig 3 38 Diagrams of the tubular systems in skeletal (*A*) and cardiac (*B*) muscle In cardiac muscle the T-tubules are larger than in skeletal muscle and the lateral sacs differ in shape and wrap partly around the T-tubule The fenestrated central cistern at the M band of skeletal muscle is absent in cardiac muscle and the branching of the longitudinal tubules is more irregular (Redrawn from Legato 1969)

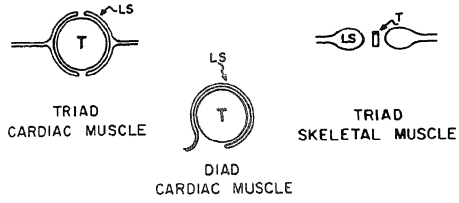


Fig 3 39 Diagrammatic concepts of triadic and diadic relationships in cardiac and skeletal muscle sarcomeres (Redrawn from Legato 1969)

diac muscle sarcoplasmic reticulum has no predictable central modification and one or more tubules often pass over the Z band into the next sarcomere Skeletal muscle sarcotubules end in a terminal dilatation, the lateral sac, whereas in the myocardium the longitudinal tubules usually divide into two arms which are wrapped around the transverse tubules The reticular tubules from adjacent sarcomeres thus flank a large central tubule forming the so-called *triad* In some cases only one lateral sarcotubule is apposed to the T vesicle forming a *diad*

The T system is thought to play a role in two important functions the conduction of excitatory electrical impulses from the cell surface membranes into the depths of the cell to the contactile system and in the distribution of substrates to the cell interior and the removal of metabolic end products The longitudinal system is thought to function in the storage, release, and uptake of calcium ions which are responsible for initiating myofibril contraction and relaxation The juxtaposition of the T system with the L system has suggested the hypothesis that the depolarizing impulse travels along the transverse tubule from the outer cell membrane, triggering release of calcium from stores within the longitudinal tubular system The calcium then diffuses into the myofibril, activating contraction Sarcoplasmic reticulum fragments are known to accumulate calcium in vitro It is thought that they do this during contraction, thus lowering the ionic calcium available to act on myosin filaments, hence leading to relaxation

In summary, the myocardial cell consists of four major functional elements a contractile element, the myofilament, an intracellular nutritive and energy producing organelle, the mitochondrion, an electrolyte transport system, the sarcoplasmic reticulum, and an organizer, the nucleus

CONTRACTILE PROTEINS

Four major contractile proteins have been extracted from myofibrils of cardiac muscle myosin, actin, tropomyosin, and troponin (Katz, 1970) There is considerable evidence that other proteins, as yet unidentified, are present in the contractile apparatus of cardiac muscle However, only the four proteins mentioned have clear roles in the contractile process Troponin may consist of two distinct proteins The approximate concentrations of these major contractile proteins in the myocardium are given in Table 3 6

Contractile proteins of the heart convert the

Table 3 6

Approximate concentrations of major contractile proteins in the myocardium expressed as grams/kilogram of net weight

Myosin	35
Actin	15
Tropomyosin	5
Troponin	2 5

Katz, 1970

chemical energy of substrate metabolism into the mechanical energy of contraction Adjustments of cardiac contractile strength result from changes in the interactions between the contractile proteins induced by physical and chemical factors The heart normally behaves as a syncytium, responding to a single stimulus with activation of all myocardial fibers In contrast to skeletal muscle, myocardial contraction cannot be altered by variations in the number of motor units activated nor by incomplete or complete summation (tetanus) dependent upon nerve impulse frequency Tension development and shortening of myocardial fibers are dependent on two factors the resting length of the sarcomeres (the Frank-Starling mechanism) and changes in the chemical environment of the proteins during contraction The length dependent changes are related to physical alterations in the positions of the contractile proteins within the sarcomere at the onset of systole Changes in the chemical environment, such as alterations in the concentration of calcium ions or catecholamines, can alter tension development in the absence of changes in sarcomere length

Myosin is the chief constituent of the thick filament of striated muscle, both cardiac and skeletal Two high-molecular weight components having different sedimentation velocities can be identified by the analytical ultracentrifugation These two components have been designated meromyosins The more rapidly sedimenting heavy meromyosin possesses both ATPase activity and the ability to bind actin Light meromyosin, the less rapidly sedimenting component, possesses neither ATPase nor actin binding activities Heavy meromyosin is the only myofibrillar protein which, under ordinary physiological conditions, is an active ATPase The ATPase activity of cardiac myosin is similar to that of red skeletal muscle myosin and both are less active than that of white skeletal muscle myosin The units of heavy meromyosin are arranged at either end of the myosin filaments and project laterally to form the cross-bridges which are visible under the electronmicroscope The rod-shaped portion of these myosin subunits consists of the lighter meromyosin which occupies the central part of the filament

Actin is the major protein of the thin filaments It exists in two states G-actin (G = globular) is the monomer which is stable in solutions of low ionic strength, and F-actin (F = fibrous) is a polymer of G-actin monomeres which is formed on addition of salts Tropomyosin makes up about 10 to 12% of the myofibrillar protein of cardiac muscle It resembles the structure of the Z line in muscle but the location of tropomyosin has not been conclusively demonstrated, although it appears to be located in the thin filaments (actin filaments) It plays a role in modulating the interaction between actin and myosin

Troponin sensitizes the interaction between actin and myosin to the effect of the calcium ion released during excitation-contraction coupling Like tropomyosin, it is believed to be associated with the thin (actin) filaments of cardiac muscle

Unpublished experiments by R E Davies and A Delluva have shown that, in the turtle heart, ATP is used for contraction Thus, the enzymatic activity of myosin is central to the conversion of chemical energy to mechanical energy during muscle contraction, since ATP is the immediate source for the chemical energy utilized in contraction Calcium ions strongly stimulate while magnesium ions inhibit the ATPase activity of cardiac myosin If both magnesium and calcium salts are present together, the inhibitory effects of the magnesium predominate, so that magnesium ions strongly inhibit the calcium activated ATPase activity of cardiac myosin

MODULATORY PROTEINS

Tropomyosin and Troponin. In vitro combinations of purified cardiac actins and myosins possess the contractile properties of myofibrils They hydrolyze ATP and can undergo physicochemical changes in vitro which probably represent the contractile process of living muscle However, the mechanisms representing the final stages in excitation-contraction coupling require the presence of tropomyosin and troponin in the actomyosin gel

The presence of troponin and tropomyosin is necessary for the sensitization of reconstituted actomyosin to the physiological actions of calcium ions The most significant action of tropomyosin is the inhibition of the interaction between actin and myosin

Troponin and Sensitization of Actomyosin to Calcium. Calcium sensitization of actomyosin requires the presence of both troponin

and tropomyosin and neither protein itself possesses this activity The sensitization of actomyosin to calcium ions appears to result from a primary interaction of the tropomyosin-troponin complex with actin in which troponin strengthens the binding of tropomyosin to F-actin Troponin has now been identified as the calcium receptive protein in cardiac muscle

In addition to the four contractile proteins mentioned in the foregoing, several other proteins and protein complexes have been prepared from muscle Two of these, α-actinin and β-actinin, appear to play a role in the structural interactions of the thin filaments and the Z line in sarcomeres, rather than modulating actin-myosin interaction Both are postulated to be involved in some way with the determination of myofilament length and sarcomere structure and possibly play a role in the production of new sarcomere units (Legato, 1970)

DIFFERENTIATION

The adult myocardium, while consisting mainly of cardiac muscle cells, is a heterogeneous tissue containing many other cell types, e g , Purkinje fibers of the specialized conduction tissue, fibroblasts, nerve fibers, phagocytes, vascular endothelium and smooth muscle cells, squamous or low cuboidal cells of the epicardial epithelial sheet, and endothelial cells of the endocardium

The developing myocardium is derived from embryonic splanchnic mesoderm and, initially, consists entirely of developing striated muscle cells Thus, it is originally an epithelial tissue, containing only developing myocytes Following development of the epicardium and endocardium, mesenchymal cells appear These nonmuscular components are probably fibroblasts derived from the endocardium or epicardium

Early cardiac myocytes contain few fibrils and large amounts of cytoplasm More fibrils form as development proceeds, the cytoplasm decreases and the fibrils become aligned in parallel pattern The term myoblast designates presumptive muscle cells as yet morphologically undifferentiated, i e , cells which have not yet acquired myofibrils Once myofibrils are present, the cells are termed myocytes The fibril to cytoplasm ratio of cardiac myocytes increases during embryogenesis until the myofibrils occupy most of the cell volume These early myocytes are epithelial cells on the basis of their ultrastructure and organization Large amounts of glandular endoplasmic reticulum are seen within the cytoplasm Such endoplasmic reticulum is largely responsible for the production of protein destined for extracellular use, and it has been suggested that these cells have a secretory or glandular function Thus, in the young heart wall, the myocardium may be described as a contractile glandular epithelium The glandular endoplasmic reticulum diminishes as the myocytes mature However, a little later, the secretory function is again evident in some ventricular myocytes in the form of dense particles within the Golgi lamellas, suggesting that this cell type may retain a secretory function throughout life (see p 3–115)

At the time of the first heart beat (in the chick embryo), well-developed fibrils containing organized myofilaments and Z bands are present

The myofibrils in early cardiac myocytes are randomly oriented Later, by an unknown mechanism, they align themselves in parallel along the same axis In well organized cells, the Z lines are at right angles to the bundles of myofibrils and the latter insert with an angle of approximately 90° into the intercalated discs These early changes are illustrated in Fig 3 40

Development of Specialized Pacemaking and Conducting Fibers. Little evidence is available on the embryological development of the specialized pacemaking and conduction tissue (sinoatrial node, atrioventricular node, bundle of His, and Purkinje network) of the adult heart (Lieberman, 1970) Morphological findings and physiological evidence (transmembrane potential characteristics) support the hypothesis that remnants of the sinus venosus and primitive atrial canal (atrioventricular ring) are the precursors of the adult sinoatrial junction, atrioventricular node, and bundle of His The Purkinje network is thought to originate from strands of primordial myocardial tissue pulled out into fascicles by the growing ventricular wall However, corroborative structure-function studies to establish the origin of the His-Purkinje system are lacking

Nonmuscular Components of Cardiac Tissue. Initially, the myocardium is a pure population of muscle cells The nonmuscular components arise from other sources The appearance of the epicardium and endocardium is associated with the entry of nonmuscular elements, such as fibroblasts, into the myocardial

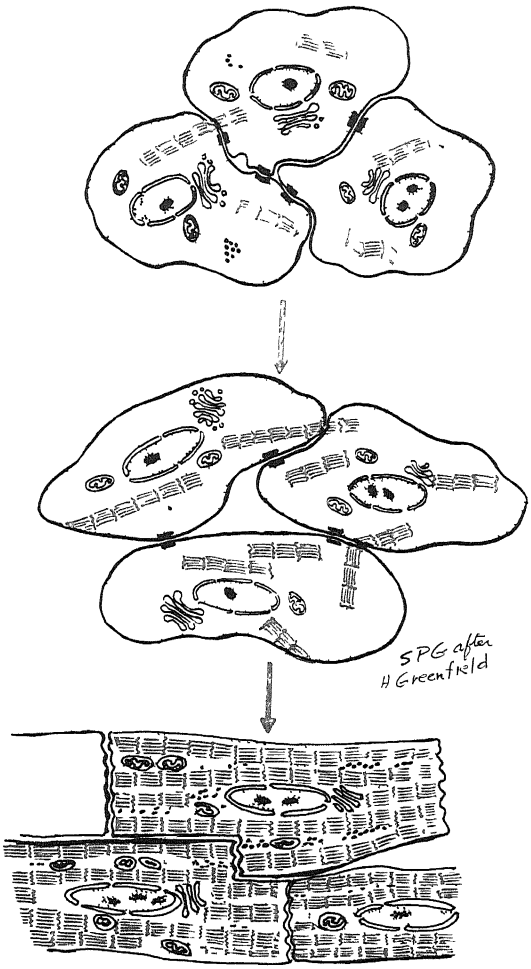


Fig 3 40 Differentiation of cardiac myocytes Three stages in early development of myocardial cells are diagrammed As maturation proceeds the cells become less rounded the number of fibrils increases and their alignment becomes more regular Then more and more myofibrils appear and intercalated discs form and are oriented at right angles to the fibrils The sarcoplasm is crowded out by the myofibrils becoming restricted to the region at the nuclear poles Mitochondria line up in rows between the fibrils and desmosomes appear at lateral cell borders (Redrawn from Manasek 1970)

tissue, apparently, they are derived from the endocardium and epicardium.

Properties of Myocardial Cells

CONDUCTIVITY

Myocardial cells possess the functional characteristic of conductivity, i e , they are capable of transmitting action potentials Although individual cells are separated by plasma membranes, the impulse can pass from cell to cell, presumably through the specialized junctions known as "tight junctions" or nexuses (q v)

CONTRACTION

The only important function of the heart is that of a muscular pump Contraction of the myocardium differs from that of skeletal muscle in several respects as follows *1*) the individual cells, while anatomically discrete, transmit impulses across the cell boundaries and thus behave as a functional syncytium A stimulus above threshold thus initiates a transmitted impulse which excites all myocardial fibers, producing an "all-or-nothing" response of the entire myocardium *2*) The active state (q v), and duration of contraction are longer *3*) The refractory period is longer and myocardial cells can neither be summated temporally nor tetanized *4*) The rate of force development during contraction is slower and velocity of shortening is slower *5*) The maximum tension developed per unit of cross-section of muscle is only $\frac{1}{3}$ to $\frac{1}{2}$ that of skeletal muscle

The physiological features distinguishing cardiac from red and white skeletal muscle are outlined in Table 3 7 Tension development in skeletal muscle is regulated by the frequency of motor nerve impulses (temporal summation, tetanus) and by variations in the number of muscle units activated during a single contraction Cardiac muscle is not organized on a muscle unit basis (nor are individual muscle fibers directly innervated), and behaves as a physiological syncytium so that all myocardial cells are activated during each contraction in an "all-or-none" fashion Further, under physiological conditions, myocardial cells cannot be summated temporally nor can they develop tetanic contraction

When compared with certain types of smooth muscle on the one hand and red and white skeletal muscle on the other, myocardial cells have structural and functional features more closely allied to those of smooth muscle and red skeletal muscle than to white skeletal muscle Smooth muscle cells are relatively poor in myofibrils, are often arranged as a syncytium and contract slowly. Similarly, red skeletal muscle is richer in sarcoplasm than white skeletal muscle and contracts more slowly than the latter Cardiac muscle, like certain types of smooth muscle, is arranged as a functional syncytium and is rich in mitochondria like red

Table 3 7
Regulation of contraction in cardiac and skeletal muscles

Mechanism	Cardiac Muscle	Red and White Skeletal Muscle
Ability to summate responses to rapidly delivered stimuli	None	Marked
Ability to vary the number of active muscle fibers	None	Marked
Ability to alter contractility	Marked	Little
Ability to alter contractile response by changing of initial fiber length	Present functions at short sarcomere lengths	Present functions at long and short sarcomere lengths

From Katz 1967

skeletal muscle The abundant mitochondria of both red skeletal muscle and cardiac muscle regenerate ATP by oxidation of fats and, to a lesser extent, carbohydrates There is much less reliance on phosphagen stores (creatine phosphate) White skeletal muscle regenerates ATP through the less efficient pathways of anaerobic glycolysis, utilizing as its major energy source creatine phosphate These various features which smooth muscle and red striated skeletal muscle share with cardiac muscle correlate with functional similarities, e g , smooth muscle, red striated muscle, and cardiac muscle contract more slowly and engage in more sustained activity than white skeletal muscle

Cardiac contraction or systole results from the aggregate shrinking of individual sarcomeres In accordance with the sliding-filament concept of Huxley, the thin actin filaments, which are anchored in the substance of the Z line at either end of the sarcomere and interdigitate with the thicker myosin filaments, move toward one another during contraction Neither the thick nor the thin filaments shorten during contraction

EXCITATION AND THE MEMBRANE SYSTEM OF CARDIAC CELLS

As in all excitable cells, excitation is mediated in myocardial cells by depolarization of cell membranes (sarcolemma) The sarcolemma in turn gives rise to the intercalated disc and the transverse tubular system The intercalated disc is made up of multiple convolutions formed of adjacent sarcolemmal membranes The transverse tubular system consists of sarcolemmal invaginations which extend from the cell surface throughout the entire thickness of the myofibril

The intercalated discs contain tight junctions of low electrical resistance, which permit rapid transmission of depolarizing impulses from cell to adjacent cell The transverse tubular network extends the extra cellular compartment to all levels of the myocardial cell, shortening diffusion distance from cell exterior to structures in the cell interior Depolarization and repolarization can occur in intimate association to intracellular structures Thus, the transverse tubular system is considered an active part of the myocardial cell membrane providing for the transmission of excitation throughout the fiber and a potential pathway for the diffusion and exchange of substances between the intra- and extracellular spaces The longitudinal sarcoplasmic tubular system, on the other hand, does not open to the cell exterior At the level of the I band and Z line the tubules and expanded ends of the intracellular longitudinal tubular system form intimate contacts as described in the foregoing

EXCITATION-CONTRACTION COUPLING

Ionic calcium appears to be the link between excitation and contraction The subcellular sarcoplasmic reticulum is known to accumulate, store, and release calcium within the muscle cell This tubular network envelops the entire circumference of the myofibril and extends throughout the sarcomere

The lateral sacs of the longitudinal sarcoplasmic tubular system are found in close association with the transverse tubules (diads and triads), along the sarcoplasmic membrane at the outer cell surface, and all along the course of the intercalated disc Electron microscopic studies have shown that the lateral sacs are storage depots for calcium It is hypothesized

that membrane depolarization in all of these areas releases calcium ions from lateral sacs, thus freeing them for interaction with the contractile proteins

In white skeletal muscle it is generally agreed that the source of calcium ions mediating excitation contraction coupling lies within the muscle cell, associated with the sarcoplasmic reticulum In the resting state, this system sequesters (binds) calcium within the cell but outside the region occupied by the contractile proteins Upon excitation, this calcium is released into the cell interior and becomes available for binding to troponin In skeletal muscle it is unlikely that an influx of calcium at the cell surface membrane during depolarization could contribute importantly to excitation-contraction coupling because the interval between excitation and contraction is too brief to permit calcium to diffuse to the troponin calcium binding sites In the myocardium, in which the cells are smaller, the development of the active state (q v) is slower and the action potential longer It is not possible to exclude calcium influx from the surface membrane as a source of calcium ions for excitation contraction coupling Very likely, net calcium fluxes maintained over a long period of time are responsible for some of the changes in myocardial contractility Further, phasic calcium movements across the sarcolemma likely cause beat-to-beat changes in intracellular calcium ion concentration responsible for excitation-contraction coupling The existence of the system of intracellular vesicles and tubules in cardiac muscle similar to the sarcoplasmic reticulum of white skeletal muscle implicates these structures in the release of calcium ions for excitation contraction coupling in the heart However, the sarcoplasmic reticulum in cardiac muscle is relatively sparse as compared with white skeletal muscle Further, less calcium is accumulated by cardiac muscle after glycerol extraction than is the case in skeletal muscle and less calcium-sequestering activity is present in cardiac microsomes than in skeletal microsomes Thus, changes in transcellular calcium flux, as well as release of calcium from the sarcoplasmic reticulum, appear to have a role in excitation contraction coupling

The amount of calcium released from the sarcoplasmic reticulum during any given contraction may determine the proportion of troponin molecules receiving calcium ion and thus govern the strength of myocardial contraction Transcellular calcium flux appears to play a role in determining contractility by this mechanism However, it is not clear whether calcium exchanges across the cell surface act during each contractile cycle or determine, on a long-term basis, the amount of membrane-bound calcium that may be released during each systole

In heart muscle, calcium deprivation does not abolish the action potential but does abolish contraction In skeletal muscle, absence of calcium prevents both electrical and mechanical response to stimulation Thus, calcium is required in skeletal muscle for action potential generation In resting skeletal muscle, almost all calcium is sequestered in the reticulum and the concentration of calcium in the sarcoplasm is almost zero In skeletal muscle, the calcium involved in contraction comes from the lateral sacs of the sarcoplasmic reticulum An influx of calcium from the extracellular medium is unimportant In heart muscle, however, part of the calcium involved in the contractile response comes from a superficial layer or from the extracellular medium while the remainder is from an intracellular site, probably the sarcoplasmic reticulum

Less calcium uptake activity is present in the cardiac microsomal (mitochondrial) fraction than in skeletal muscle microsomes However, this capacity is more than sufficient to consider mitochondria as a source for adequate calcium release to saturate all calcium-binding sites that control the actin-myosin interaction Whether the calcium affinity of cardiac microsomes is sufficient to bind calcium rapidly enough to account for the rapid calcium removal causing relaxation is uncertain It is possible that mitochondria, which have been shown to have a capacity to take up calcium, may play a role in the relaxation process However, the evidence that uptake and release of calcium by myocardial mitochondria play a significant role in cardiac excitation-contraction coupling is not conclusive

How calcium ions cause sarcomeric shortening is somewhat speculative Katz (1970), has proposed a model which defines the interactions between the four proteins myosin, actin, tropomyosin, and troponin (Fig 3 41) The interaction between actin and myosin is regulated by the amount of ionic calcium made available for binding to troponin The calcium

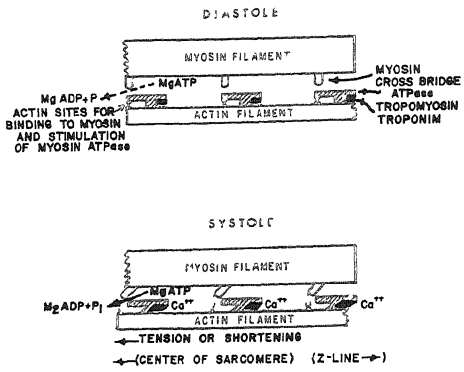
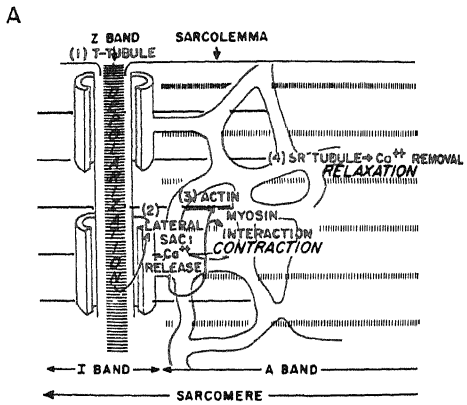


Fig 3 41 Possible interactions between the two contractile proteins actin and myosin and the two modulatory proteins tropomyosin and troponin *Upper* in the resting state (diastole) tropomyosin is depicted as the interaction inhibitor *Lower* during systole addition of calcium abolishes inhibition of the actin myosin interaction by the tropomyosin-troponin complex This appears to result from binding of calcium to troponin and an action of troponin and tropomyosin together Release of inhibitory action of modulatory proteins permits the ATPase stimulating and physicochemical interactions between actin and myosin to proceed The active state is established and the muscle shortens Maximal shortening velocity (V_{max}) appears to be governed by the rate at which actin-activated myosin cross-bridges hydrolyze ATP Maximum tension (P_0) seems to be determined by the number of interacting sites between actin and myosin (Redrawn from Katz 1970)

acts to release an inhibitory action of the tropomyosin-troponin complex upon reactive sites on the actin molecule The latter is then able to interact with myosin It is assumed that the initial action of calcium is to induce a conformational change in the troponin molecule This, in turn, leads to a reversal of the inhibitory action of tropomyosin, possibly through a troponin-tropomyosin interaction Thus, calcium permits release of the inhibitory action of the modulatory proteins, allowing the ATPase-stimulating and physicochemical interactions between actin and myosin to occur These events establish the active contractile state required for muscle shortening This speculative explanation is schematized in Figure 3 41

The contractile event depends on the presence of calcium and relaxation upon its removal When calcium ions are removed from the area of the myofilaments, cross-bridge formation between actin and myosin ceases and the sarcomere returns to its resting length Thus, the sarcoplasmic reticulum with its ability to bind, transport, and resequester calcium in an inactive site (lateral sac) is considered the sub-cellular system which produces contraction, regulates tension, and achieves relaxation These various functions are summarized in Figure 3 42*A* and *B*

In summary, it is thought that the depolarizing impulse travels along the sarcolemma and is propagated throughout the cell via the T tubular network Depolarization of the membranes of the lateral sacs associated with T tubules and those beneath the sarcolemma results in calcium ion release Calcium ions induce interaction between actin and myosin and contraction results Relaxation is the result of



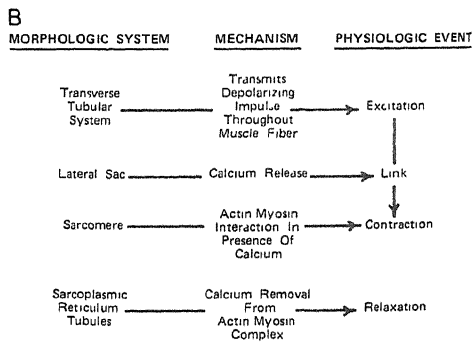


Fig 3 42 *A* Subcellular systems involved in excitation-contraction coupling and relaxation The depolarizing impulse travels over the sarcomeric membrane and along the T-tubule membrane into the depths of the cell Depolarization causes calcium ion release from storage depots of the lateral sacs Calcium ions bind with troponin of the tropomyosin-troponin complex and somehow abolish the inhibitory effect of these modulatory proteins on the interaction between actin and myosin and contraction occurs Relaxation occurs when this freed calcium is removed by being taken up again by the lateral sacs ***B***, the events of excitation contraction coupling are summarized in this table (Redrawn and modified from Legato 1969)

sarcotubular removal of calcium from the area of the myofilaments

ACTIVE STATE

The active state represents a capacity of the contractile element to shorten and do work but not the transformation of this into external mechanical performance (Holland and Sabatini-Smith, 1969) The mechanical output of a strip of cardiac muscle during a single contraction is only an indirect representation of the degree of activity of the contractile component itself The external manifestation of activity is dependent on the stretching of the series component and is delayed Hill used the term "active state" to designate the degree of activity of the contractile unit It may be expressed as a maximal isometric tension that the contractile unit is able to develop or just bear without lengthening at a given moment To measure the active state, recording must be arranged such that no movement of the series elastic element occurs Then the capacity to shorten and do work is translated into mechanical performance Study of the active state in cardiac muscle depends on indirect methods such as the "quick release" method of Jewell and Wilkie (1958) The shortening velocity of the contractile unit at various moments after stimulation can be determined by this method The speed of shortening at any given moment is proportional to the ability of the contractile unit to develop tension

The active state is much longer in the heart than in the skeletal muscle In rabbit papillary muscle (at a contraction frequency of 30 beats/min at 30° C), the intensity of the active state is maximal about 60 msec after stimulation, and activity begins to decline at about 100 msec The active state has diminished to approximately 80% of the peak value at the time of the maximal isometric tension (Holland and Sabatini-Smith, 1969)

An inotropic change of the myocardium may be achieved by an alteration of either two parameters of the active state its intensity or its duration A reduction of the duration of the active state will decrease the height of the contraction by giving the contractile unit less time to stress the series elastic component Thus, decreased or increased duration of active state results in reduction or lengthening, respectively, of the time to peak tension A change in intensity of the active state will produce a symmetrical alteration in the height of the isometric contraction curve

METABOLISM AND ENERGETICS

Under normal circumstances, myocardial cells operate almost exclusively on an aerobic metabolic system which provides the constant supply of high-energy phosphate bonds for mechanical and chemical work

In Table 3 8 are listed various metabolic substrates together with their percentage contribution to myocardial oxygen usage Under normal conditions the major fuel for cardiac metabolism is free fatty acids Glucose and lactate are also important contributors while amino acids, ketones, and pyruvate make lesser contributions

Olson and Piatnek (1959) divided myocardial metabolic processes into three phases (see Fig 3 43) energy liberation, energy conservation, and energy utilization

1) *Energy liberation* or production is the conversion of the energy of substrates (e g , fatty acids, glucose, lactate) into the energy of the high-energy phosphate compounds, adenosine triphosphate (ATP) and creatine phosphate (CP)

2) *Energy conservation* or storage is the steady state concentration of ATP and CP in the heart

3) *Energy utilization* is the rate and efficiency of conversion of this chemical energy to the mechanical energy (work) of cardiac contraction The major intermediary metabolic pathways of myocardial cells are diagrammed in Figure 3 44 Note that the mitochondria (sarcosomes) are the sites of the major aerobic or oxidative processes upon which the heart

Table 3 8
Percentage contribution of various metabolic substrates to total myocardial oxygen usage

Carbohydrate		Noncarbohydrate	
Glucose	17 90%	Fatty acids	67 0%
Pyruvate	0 54%	Amino acids	5 6%
Lactate	16 46%	Ketones	4 3%
Total	34 90%[a]	Total	76 9%[a]

[a] The total is more than 100% because some of the substrates removed from the blood, in amount measured by the arteriovenous differences, were stored by the heart or not completely metabolized

From Bing, 1955

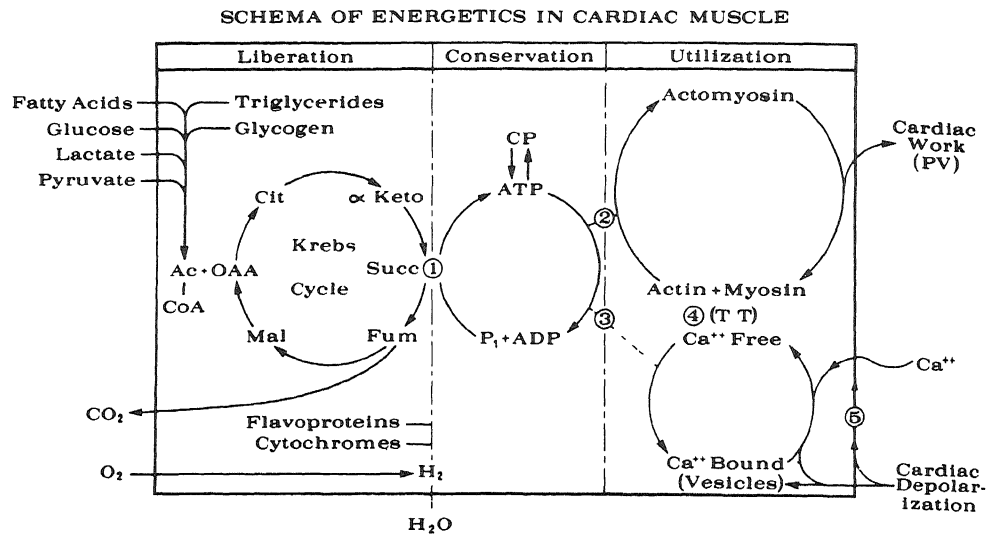


Figure 3 43 The three phases of myocardial energetics Energy coupling points affected by hypoxia include (1) oxidative phosphorylation (2) actomyosin ATPase (3) ATP energized calcium pump of the reticulum (4) troponin tropomyosin-calcium system for regulating contractility and (5) ion pumps in the cell membrane Cit citrate α Keto alpha ketoglutarate Succ succinate Fum fumarate Mal malate Ac CoA acetyl CoA OAA oxalacetate CP creatine phosphate ATP adenine triphosphate ADP adenosine diphosphate (Olson *et al* 1971)

depends primarily for energy Myocardial cells differ from white skeletal muscle cells in having a much higher concentration of mitochondria In some myocardial cells the mass of mitochondria equals that of the myofibrils

In white skeletal muscle the major metabolic pathways are anaerobic while in red skeletal and cardiac muscle they are aerobic, thus accounting for the high concentration of mitochondria in the latter muscle types

In cardiac muscle during the energy liberation phase, the major myocardial substrates (see Table 3 8) are metabolized to the two-carbon molecule, acetyl-coenzyme A, which enters the citric acid (Krebs) cycle in the mitochondria Then follows the step-wise release of carbon-hydrogen bond energy

In the second, or energy conservation phase, the energy release through hydrogen or its electron is transferred into the bonding energy of the high-energy phosphate bond (~P) through oxidative phosphorylation This process occurs in the mitochondria also The result of oxidative phosphorylation in the mitochondria is the formation of creatine phosphate (CP), adenosine monophosphate (AMP), adenosine diphosphate (ADP), and adenosine triphosphate (ATP)

The third, or energy utilization phase, occurs when ATP is hydrolyzed by adenosine triphosphatase to ADP and inorganic phosphate (P_i) at the time of muscular contraction ATP is the immediate source of energy for muscular contraction and other chemical cellular processes such as modifying substrates, producing glycogen, synthesizing lipids and proteins, and maintaining ionic concentration gradients across the cell membrane The relationships between these biochemical events, myocardial contraction, and cellular maintenance are diagrammed in Figure 3 45 The physicochemical interaction of actin and myosin utilizes ATP The latter is then replenished from high phosphate bonds from creatine phosphate (CP) or from the electron transport chain The increased amounts of adenosine diphosphate (ADP) and inorganic phosphate (P_i) resulting from muscular contraction act to stimulate oxidative phosphorylation, the citric acid cycle and glycolysis This increases high-energy phosphate bond production High levels of ATP depress these processes Thus, the relative levels of ADP and P_i on the one hand and ATP on the other serve as a regulatory feed-back control governing aerobic mitochondrial processes leading to the production of high-energy compounds Creatine phosphate (CP) functions as a small but immediately available source of high-energy phosphate bonds

To understand the dependence of cardiac

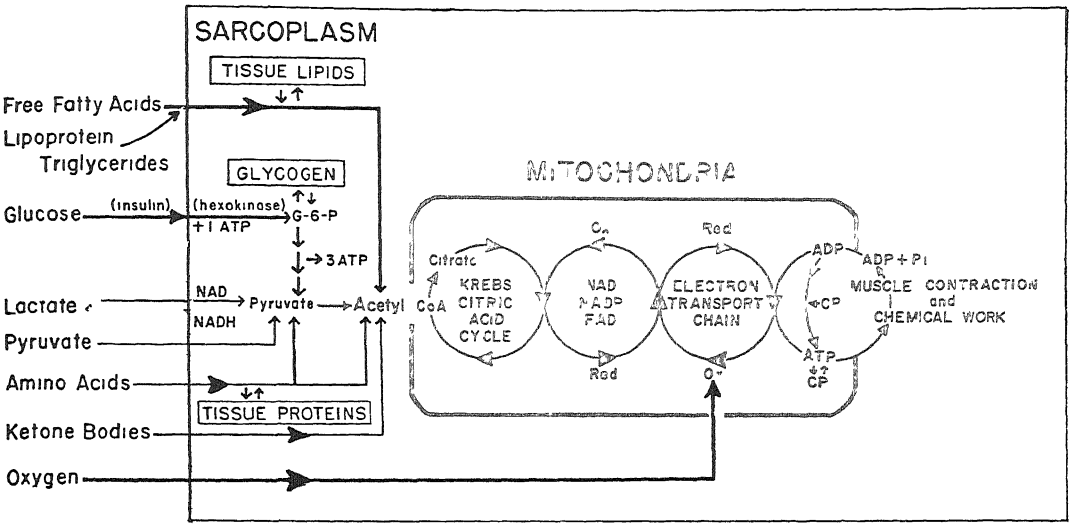


Fig 3 44 Schematic diagram of a myocardial cell illustrating the major pathways of cardiac metabolism Most substrates are taken up more or less in proportion to their arterial concentration above their extraction threshold The major reactions of glycolysis which are anerobic take place in the sarcoplasm (cytoplasm) The mitochondria are responsible for the major aerobic or oxidative processes including the oxidation of free fatty acids (FFA) the citric acid cycle which utilizes acetyl Coenzyme A (acetyl CoA) from many sources and the complex respiratory chain or electron transport chain (ETC) which oxidizes reduced flavin and adenine coenzymes formed in the citric acid cycle The process of oxidative phosphorylation results in the formulation of adenosine triphosphate (ATP) and creatine phosphate (CP) ATP is the immediate source of energy for muscular contraction and for many chemical reactions while CP serves as a reservoir of high energy phosphate bonds FFA are metabolized predominantly in the mitochondria to acetyl CoA which enters the citric acid cycle The contribution of lipoprotein triglycerides to the fatty acids metabolized by the heart is uncertain Glucose is transported across the cell membrane under the control of insulin and is then phosphorlyated in the cytoplasm by hexokinase to glucose-6-phosphate (G-6-P) a reaction which utilizes one high energy bond from ATP In the subsequent steps in the metabolism of G-6-P to pyruvate each molecule of G-6-P yields 4 high energy phosphate bonds however one of these is used in the important and rate-limiting phos phorylation of fructose-6-phosphate to fructose-1 6-diphosphate by phosphofructokinase Lactate is normally taken up and converted to pyruvate which is then converted to acetyl-CoA to enter the citric acid cycle The enzyme phosphorylase which exists in two forms catalyses the reaction glycogen + $P_i \rightarrow$glucose-1-phosphate which is then changed to G-6-P to enter the glycolytic pathway Hypoxia inhibits many mitochondrial functions including the ETC and the citric acid cycle, with a resultant decrease in CP and ATP Hypoxia also stimulates glycogenolysis glycolysis and uptake of glucose Excess pyruvate produced during hypoxia is released by the heart as lactate The metabolism of large amounts of FFA or ketone bodies increases the intracellular concentration of citrate which inhibits phosphofructokinase and therefore glycolysis and the uptake of glucose Abbreviations NAD nicotinamide-adenine dinucleotide NADH reduced nicotinamide-adenine dinucleotide, NADP nicotinamide-adenine dinucleotide phosphate FAD flavin-adenine dinucleotide Red reduction Ox oxidation (From Schlant 1970)

muscle on the continual oxygen supply, it is useful to compare the patterns of energy production and utilization with those of red and white skeletal muscle (Katz, 1970) The heart, like red skeletal muscle is organized to maintain its work over long periods of time without rest, both of these types of muscle rely upon oxidative metabolism for ATP regeneration and have little ability to function anaerobically or to incur oxygen debt This is in contrast to white skeletal muscle which is organized for brief bursts of intense activity followed by periods of relaxation The substrates, metabolites, pathways of intermediary metabolism, and intracellular energy reserves of the two types of muscle are strikingly different (Table 3 9) Cardiac muscle derives its energy mainly from oxidative metabolism It contains large numbers of mitochondria which are rich in the enzymes of the citric acid cycle and oxidative phosphorylation The myoglobin content is large, favoring oxygen diffusion into the fiber In white skeletal muscle the chief source of energy comes from the metabolism of carbohydrates (glycogen, glucose) During activity, these substrates are not oxidized to carbon dioxide and water Rather, lactic acid is produced and an oxygen debt develops The enzymes catalyzing aerobic glucose utilization and anaerobic glycolysis are more abundant in

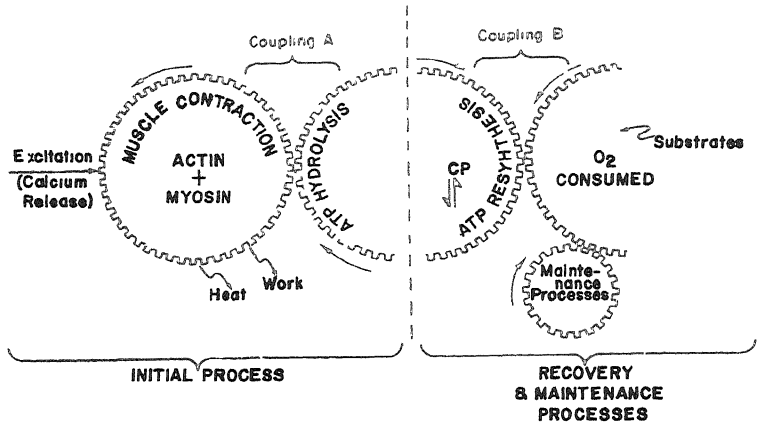


Fig 3 45 Diagram of mechanochemical coupling in cardiac muscle The interaction of actin and myosin during contraction is initiated by Ca^{++} release (left hand panel initial process) and utilizes ATP through Coupling A Oxidative phosphorlyation provides for the formation of creatine phosphate (CP) and adenosine triphosphate (ATP) CP serving as a reservoir for high energy phosphate bonds available for ATP resynthesis Oxygen utilization is related to CP and ATP synthesis through Coupling B Oxygen is also consumed in various cellular maintenance processes (Redrawn from Braunwald et al 1967)

Table 3 9
Energy sources for contraction of red and white skeletal and cardiac muscle

Energy Source	White Skeletal Muscle	Red Skeletal and Cardiac Muscle
Major substrates	Carbohydrate	Lipid
Major metabolic pathways	Anaerobic	Aerobic
Major products of metabolism	Lactic acid	CO_2 and H_2O
Oxygen debt on activity	Significant	Minor
Reserve of phosphate bond energy (phosphocreatine)	Significant	Minor
Ability to function in the absence of oxygen	Good	Poor

From Katz, 1967

white skeletal muscle Cardiac muscle is richer in enzymes providing for the oxidative metabolism of pyruvate, succinate, and acetoacetate The heart is specialized metabolically to function aerobically, requiring a continual oxygen supply and large numbers of mitochondria White skeletal muscle can contract in the absence of oxygen, utilizing anaerobic carbohydrate metabolism and its large phosphocreatine reserve However, it needs a recovery period during which its glycogen and creatine-phosphate reserves are replenished by oxidation of lactic acid. Only under anaerobic conditions does the heart utilize glycolysis and deplete its slender creatinephosphate reserve Thus, the heart, like red skeletal muscle, depends upon oxidative metabolism for ATP regeneration and requires a steady supply of oxygen

The energy utilization phase of cardiac and red skeletal muscle also differs characteristically from white skeletal muscle The ATPase activity of myosins from the former is less active than that of white skeletal myosin The maximal shortening velocities of both red skeletal and cardiac muscles are slower than that of white skeletal muscle, presumably because shortening velocity may be determined by the rate of ATP hydrolysis by the contractile proteins. The relatively lower tensions developed by cardiac muscle, as contrasted with white

skeletal muscle, may be attributed in part to the larger number of mitochondria in myocardial cells These, by taking up considerable intracellular space, reduce the amount of contractile protein present per unit volume of myocardial tissue

REGENERATION

There is no unequivocal evidence that adult myocardial cells can divide (Grove et al , 1969) This is in contrast to the differentiating embryonic cells which can multiply by cellular division Embryonic and early postnatal growth of the heart is accomplished by hyperplasia of myocardial cells It appears that any subsequent myocardial enlargement can only be accomplished by hypertrophy of the individual myocardial cells through enlargement and multiplication of intracellular structures (mitochondria, myofilaments) This absence of the capacity for hyperplasia and regeneration remains somewhat controversial since there are indirect evidences that the number of myocardial cells may increase during hypertrophy in the adult individual (Linzbach, 1960, Beznak, 1969) From tissue culture studies of myocardial cells it is known that, in contrast to skeletal muscle, both contractile and undifferentiated cells can divide In developing skeletal muscle, once the cell begins contractile activity, mitotic cell divisions and DNA synthesis cease Whether or not undifferentiated stem cells (myoblasts) may remain in the myocardium and form new cells under appropriate stimuli, and whether hyperplasia through cell division or activation of latent myoblasts can occur in early life (e g , hypertrophy accompanying congenital heart disease) is unknown However, at present the concensus is that ordinarily an increase in myocardial mass can be accomplished in the adult only through enlargement of cardiac cells already formed

DILATATION

The term cardiac dilatation refers to increased capacity of the heart chambers Alterations in cardiac volume occur within physiological limits as part of the mechanism for changing cardiac output Enlargement of cardiac chambers beyond physiological limits occurs both acutely and chronically in various abnormal states

Three factors are involved in acute dilatation sarcomere lengthening, "fiber slippage," and stratification change Two of these factors are represented schematically in Figure 3 46 In acute dilatation, sarcomeres longer then 2 2 μ appear, and in electron micrographs, an H zone (Fig 3 36 page 3-29), normally absent in cardiac sarcomeres, can be seen (Fig 3 46*B*) "Fiber

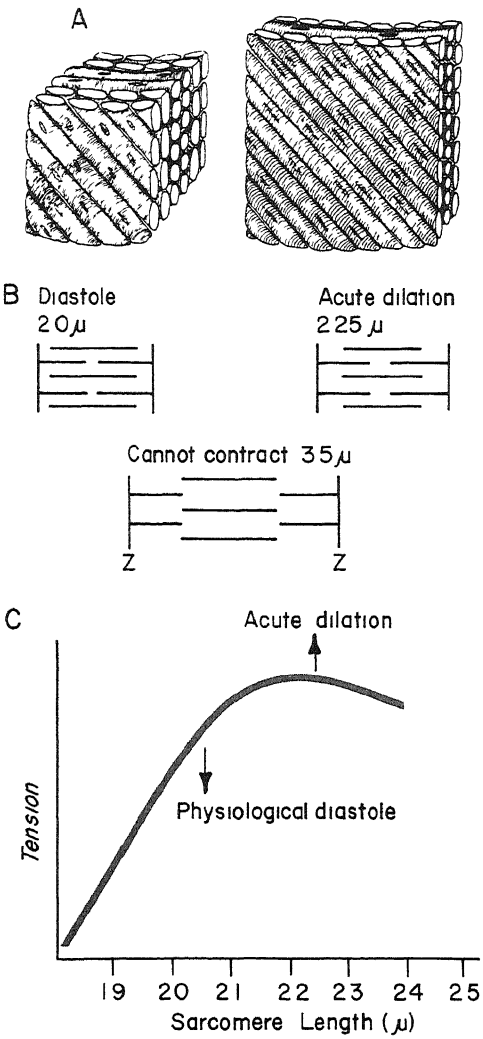


Fig 3 46 ***A,*** schemata of stratification change The same part of the right ventricular wall is represented in normal state and with marked dilatation ***B,*** schematic representation of sarcomeres in different states of stretch showing the relations of the thin and thick filaments ***C,*** Length tension diagram of an isolated papillary muscle showing the sarcomere lengths for physiological diastole and acute dilatation (Redrawn from Hort 1968)

slippage" means that individual fibers slide past one another in different directions longitudinally A change in stratification of fibers is believed to account for dilatation in a direction at right angles to myocardial fibers In this process there is a rearrangement of fibers from adjacent layers of myocardium As the myocardium is stretched, the individual fibers become thinner and fibers from adjacent layers slip between one another This increases the number of fibers within a single layer and decreases the number of layers across the myocardial wall In this way, wall thickness is reduced and surface area increased The term *plastic dilatation* has been applied to longitudinal "slippage" and lateral rearrangement of myocardial fibers

Dilatation, which increases sarcomere length and increases the internal diameter of cardiac chambers, has two functional implications Increasing sarcomere length, within limits, increases the contractile strength of myocardial fibers (Frank-Starling relationship, pp 3-88 and 3-119) An increase in chamber diameter increases the wall tension for a given intracardiac pressure in accordance with the Laplace relation (Tension = Pressure $\times$ π radius^2, see page 3-92 and Fig 3 96) Physiologically, increased stretch results in increased wall tension but, at the same time, force of contraction is increased However, if the sarcomeres are stretched too far, contractile strength decreases and the heart still must operate at an increased wall tension for a given level of intracardiac pressure Studies on experimental animals in which acute cardiac dilatation has been induced (e g , acute asphyxia) indicate that sarcomere stretch beyond physiological limits does not ordinarily occur (Fig 3 46) The increase in sarcomere length which does take place cannot account fully for the degree of observed dilatation, so that fiber "slippage" and stratification change must play a significant role It does not appear that the sarcomeres become stretched to the point that this factor alone can account for contractile weakness nor that individual sarcomeres are stretched to the point that they cannot contract at all However, sarcomere length is increased to the point that there is little reserve for increase in contractile strength through the Frank-Starling mechanism

In chronic dilatation it has been shown experimentally that fiber "slippage" and stratification change as well as moderate increase in sarcomere length occur, as in acute dilatation However, in chronically dilated human hearts, Linzbach and Kyrieleis (1968), did not find any increase in sarcomere length Additionally, there is the possibility (unproven) that new sarcomeres may be synthesized in series to lengthen cardiac fibers Dehiscences or separations of fibers at intercalated discs have been found by electron microscopy in tissue from the auricular appendages of chronically dilated human hearts (Poche, 1969), and increased number of sarcomeres per fiber have been identified in hypertrophied hearts (Laks et al , 1969)

In summary, acute and chronic dilatation result from "slippage" and stratification change of fibers primarily Sarcomere length is increased moderately, at most, or, as some studies have shown, not at all Because of the increased diameter, the heart must operate at increased wall tension for a given intracardiac pressure Insofar as sarcomere lengthening occurs, the reserve for increased contractile force with increased sarcomere length is compromised (Linzbach and Kyrieleis, 1968, Hort, 1968)

HYPERTROPHY

When the heart is subject to mechanical overload, either through disease states, long-term heavy physical work, or experimental intervention, increase in myocardial mass results The normal heart in human adults ranges from approximately 250 to 350 g in weight Those in athletes or laborers may reach 500 g , but in diseased hearts weights of 500 to 1000 g or even more are seen The principal, or perhaps the exclusive, adaptive change occurring in these conditions is hypertrophy of the individual myocardial fibers, accompanied by an increase in other tissue elements (connective tissue, vasculature) It has been found through light and electron microscopic studies of hypertrophied cardiac tissue that the myocardial cells are increased in diameter and length For example, the average diameter of the normal human heart myocardial cell is 15 μ while in hypertrophied hearts fibers from 25 to 30 μ are found The increase in cell length is the result of adding more sarcomeres to the cell rather then the result of stretching of the sarcomeres (Laks et al , 1969) The locus for addition of sarcomeres may be at the intercalated discs, but this speculation requires further evidence

Legato (1970) has provided electronmicroscopic evidence suggesting that the Z line substance is essential for new sarcomere production In myocardial tissue from adult human hearts, newborn puppy hearts, and in rat ventricular cells grown in tissue culture proliferations of the Z line were observed In certain of the masses of Z substance, some of which had reached a width of 1 5 to 2 0 μ, filaments embedded in an amorphous matrix were observed These findings led to the postulate that the Z substance may serve as a source or template for the production of new sarcomeres in the developing or enlarging cardiac cell The increase in diameter is accomplished by the addition of more myofibrils The ratio of the volume of mitochondria to the volume of myofibrils increases slightly at first, but later returns to normal (Meerson, 1969)

With normal growth of the heart during maturation, the individual myocardial cells do not multiply but become thicker and longer In infants there is one capillary for every four or five myocardial fibers, while in the adult heart the ratio is 1 1 (Linzbach, 1960) In the presence of hypertrophy the adult ratio remains The idea that *capillary inadequacy* is a cause of hypoxia in hypertrophied hearts does not appear valid In marked hypertrophy (heart weights exceeding 500 g) the ratio between coronary artery and coronary ostia diameters decreases This may restrict myocardial perfusion in hypertrophy

Number of Cells There is no crucial evidence for an increase in numbers of myocardial cells, although some investigators suggest that this too may occur, especially in extreme myocardial hypertrophy (Beznak, 1969, Linzbach, 1960) Myocardial cells which have been induced to hypertrophy by experimentally produced mechanical cardiac overload are known to return to their original size when the cardiac workload is returned to normal (Beznak, 1969) Further, if the work of the heart is reduced below normal, myocardial cells will grow smaller (atrophy) Clearly, the myocardial cell possesses the ability to increase or decrease myofibrillar mass in response to altered workload This, in turn, must be accomplished by alterations in protein synthesis which are somehow linked to changes in the level of contractile function In Figure 3 47 each line schematically represents various interrelations between basic myocardial processes The relations

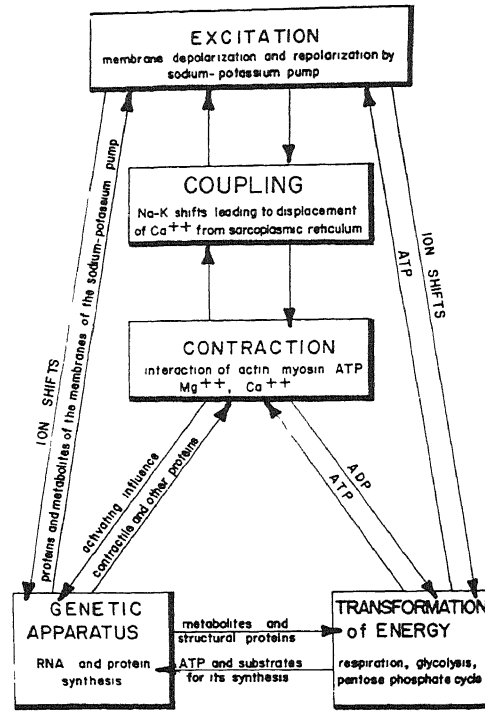


Fig 3 47 The interrelations of various fundamental myocardial processes (Redrawn from Meerson 1969)

between excitation, excitation-contraction coupling, contraction, and energy transformation have already been discussed To these fundamental myocardial cellular processes must be added that of protein synthesis, which is controlled by the genetic apparatus of the nucleus The possible regulatory mechanisms implied as affecting the genetic apparatus are hypothetical

As with other living cells, cardiac fibers maintain a state of dynamic equilibrium between synthesis and breakdown of cellular elements For example, myofibrillar and mitochondrial proteins of rat heart muscle turn over with half-lives of ten to twelve, and five to six days, respectively (Rabinowitz and Zak, 1972) The other intracellular organelles and membranes (e g , sarcoplasmic reticulum) are maintained by continual replacement or renewal of structural proteins This process is under the control of the genetic apparatus in the nucleus (and also in part in the mitochondria) and involves nuclear DNA and the sarcoplasmic RNA system Thus, the cellular structures are maintained at a given level, dependent on the balance struck between the

forces of synthesis and destruction In the case of cellular hypertrophy involving marked increase in contractile proteins, synthesis must be increased so that contractile protein production exceeds loss For this to occur, the genetic apparatus must respond appropriately to variations in the demand on cells for contractile function, and to changes in the rate of breakdown of cellular structures (so-called "wear and tear" of cell structures) The mechanism linking the DNA-RNA system to cardiac activity is obscure However, for hypertrophy to follow increased functional demand on the heart, it is clear that a mechanism must exist which regulates protein synthesis in accordance with contractile function Thus, intensification of the contractile process not only steps up metabolic energy transformations but also has an activating influence on the level of activity of the genetic apparatus

In Figure 3 48, a speculative schema on the pathogenesis of cardiac hypertrophy is illustrated

Workload Cardiac hypertrophy occurs when the workload of the heart is increased abnormally The heart responds by what Meerson (1969) terms compensatory hyperfunction Physiologically, two different types of compensatory hyperfunction are postulated isotonic hyperfunction and isometric hyperfunction Isotonic hyperfunction results when there is an increased blood inflow to the heart as in valvular insufficiency, arteriovenous fistulas, anemia, etc Isometric hyperfunction occurs in valvular stenosis or systemic or pulmonary hypertension Clinically, hypertrophy is also found in the hearts of patients with myocardial or coronary artery disease, accompanied by cardiac dilatation In these cases, a primary increase in workload has not occurred However, wall tension has increased in accordance with the Laplace relationship (q v) Accordingly, a greater radius of a cardiac chamber resulting from dilatation will increase contractile tension and provide a stimulus to hypertrophy Badeer (1968) points out that in all three conditions the mechanical stimuli for hypertrophy require an increased energy expenditure per unit mass of myocardium per beat In Figure 3 48, increased contractile stress per unit cross-sectional area is depicted as an initial change resulting in augmented energy expenditure per beat and stimulation of the genetic DNA-RNA system Hypertrophy results and, as represented in the left hand series of boxes, may be followed by compensation in which a steady state is produced with normal energy expenditure and maintenance of a normal concentration of contractile elements This is considered a stable state of

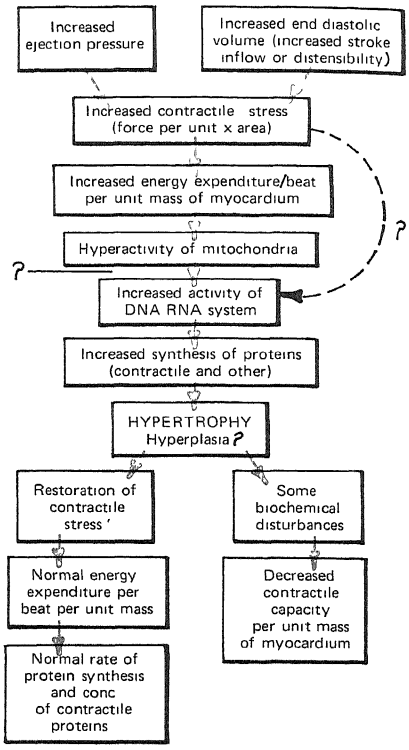


Fig 3 48 The pathogenesis of cardiac hypertrophy (Redrawn from Badeer 1968)

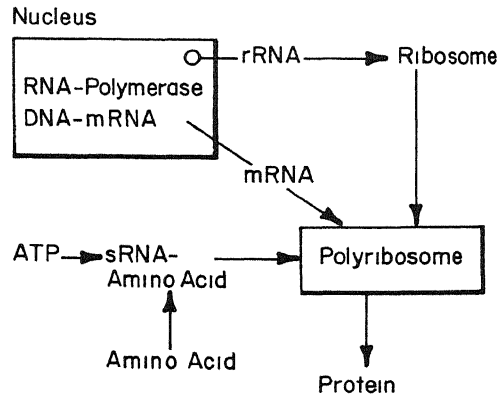


Fig 3 49 Schema of protein synthesis regulation Hemodynamic overload of the heart stimulates nuclear and nucleolar RNA synthesis initiating the hypertrophic process (Redrawn from Scheuer, 1970)

hypertrophy and hyperfunction However, as shown on the right hand side, hypertrophy may fail to restore the functional capacity of the myocardium so that its contractile performance per unit mass may be decreased Decrease in maximum shortening velocity and tension development has been demonstrated in myocardial tissue from non-failing hypertrophied hearts (Spann et al , 1967) This will ultimately lead to congestive heart failure

Protein Synthesis The mechanism of protein synthesis is shown in Figure 3 49 This is under control of the DNA-RNA system and involves three types of RNA ribosomal RNA (rRNA), messenger (mRNA), and transferring (soluble or sRNA) In accordance with present views, ribonucleic acid (RNA) formation occurs in the nucleus and protein synthesis in the ribosomes DNA-dependent messenger RNA (mRNA) synthesis is controlled by the enzyme RNA polymerase Messenger RNA transfers the genetic information from the nuclear DNA to the ribosomes and RNA acts as a template for the synthesis of cellular proteins from amino acids An aggregate of ribosomes, polyribosome, assembles amino acids linked to sRNA into proteins Ribosomal RNA is repeatedly programmed by messenger RNA from the nucleus The amino acids are activated by ATP and transferred to the ribosomes by transferring RNA The ribosomes contain the necessary enzymes, ions, and high energy phosphate for this synthetic process Thus, nuclear DNA regulates protein synthesis through its control of RNA manufacture In myocardial hypertrophy, increased synthetic rates for mRNA, nuclear RNA, mitochondrial RNA, and microsomal, mitochondrial, and cytoplasmic protein occur Increased amino acid incorporation into protein is found as early as three hours after a pressure load is imposed on the isolated heart In microsomal protein, synthetic rates increase as early as 1 hour after an increased cardiac load is applied Various hormones which support protein synthesis play a role in cardiac hypertrophy Growth hormone must be administered in hypophysectomized animals with aortic constriction if hypertrophy is to occur Thyroid stimulating hormone potentiates the effect of growth hormone Adrenalectomy decreases the heart weight in normal animals but does not prevent hypertrophy with an increased cardiac load Insulin and male hormones also influence the hypertrophic response to mechanical overload However, since increased protein synthesis can be demonstrated in overloaded isolated hearts, these hormonal influences are not required for the initiation of the genetic process In rats, digitalis glycosides have been shown capable of protecting against the development of ventricular hypertrophy caused by chronic systolic overload induced by aortic constriction (Greenspan and Edmonds, 1969) Possibly this effect occurs because the drug prevents ventricular dilatation with the increase in intraventricular pressure

Transmembrane Potential (TMP)

The studies of Weidmann (1951), and Woodbury et al (1951) showed that the potential difference across a limited area of muscle membrane of a single heart fiber could be faithfully recorded in the beating heart by use of an intracellular microelectrode of the type developed by Ling and Gerard (1949) Glass tubes of about 1 mm external diameter are drawn out to an extremely small tip of about 0 2 μ When these are filled with concentrated potassium solution and inserted through the fiber membrane into the sarcoplasm, they can be used to measure the potential difference between the outside and inside of a single fiber in the beating heart By cutting off the tip of the glass capillary electrode and suspending it on a wire, the electrode is free to move up and down Thus artefacts from heart movements are minimized and recordings from human hearts during cardiac surgery have been made Human atrial and ventricular muscles, excised from patients undergoing open-heart surgery, were also studied in vitro Where studied simultaneously, the transmembrane potentials from fibers in the intact heart are qualitatively and quantitatively similar to those recorded from single fibers of isolated preparations of either atria or ventricles Information thus obtained concerning the state of polarization of the muscle membrane serves as a much better basis for correlation with studies of excitability and refractoriness than the less direct and more variable measurements obtained with surface electrode techniques (Fig 3.50)

TYPICAL POTENTIALS

Transmembrane Potentials (TMP) of Ventricle and Atrium. Figure 3 51*A* shows the potential changes that are observed when a capillary microelectrode is inserted into a single fiber of a rhythmically beating ventricle When both the microelectrode and the indifferent electrode are outside the cell, there is no potential difference When, however, the tip of the microelectrode penetrates the membrane of a single myocardial fiber, a sudden potential of 80 to 90 mv is recorded with the inside of the fiber negative to the indifferent electrode This resting potential difference across the muscle membrane remains at the same value as long as the tissue is quiescent Following stimulation, however, there is an extremely rapid phase of depolarization (1 to 2 msec) and change in the membrane potential so that the intracellular microelectrode momentarily becomes positive with respect to the indifferent electrode Therefore, at the peak of the action potential, the membrane is not only depolarized but actually exhibits a reversal of polarity or exhibits overshooting amounting to 15 to 30 mv , and lasting 6 to 15 msec Immediately after this reversal there is a plateau of about 100 msec and then a moderately rapid repolarization of the membrane for 100 to 150 msec , as the resting potential is restored On pulling back the electrode, the potential jumps back to the zero level It is believed that use of an intracellular microelectrode gives a true picture of the time course of depolarization and repolarization in a single cardiac muscle fiber and provides an accurate measurement of the size of transmembrane potentials

Transmembrane potentials recorded from the atria and auricles in the open-chest dog (or from single fibers of isolated preparations) are not significantly different from those of the ventricles except that following depolarization, the atrial records show an immediate slow phase of repolarization without an intervening plateau such as occurs in ventricular muscle (Fig 3 51*B*)

The temporal relationship between the action potential of a single fiber and the local ventricular electrogram of the intact and in situ heart is shown in Figure 3 51*A* The upstroke of the action potential coincides with the R wave of the electrogram, the plateau with the R-T segment, and the phase of repolarization with the T wave The temporal relation between the electrical and mechanical activity is very close

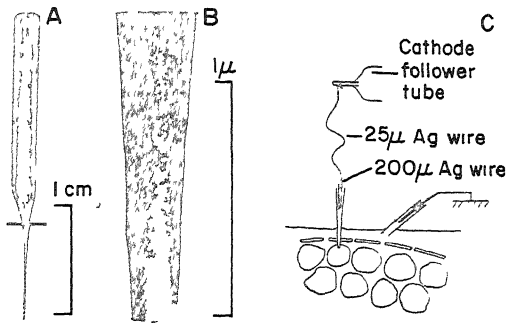


Fig 3 50 Microelectrodes used for determining transmembrane potential in single myocardial fiber as modified by Alexander and Nastuk from Ling and Gerard (1949)

(Fig 3 51*E*) Contraction starts a few milliseconds after depolarization, and tension reaches a peak when the membrane undergoes repolarization The state of contraction is maintained as long as the fiber is depolarized, and relaxation sets in as soon as the resting potential is re-established Under some conditions it has been possible to obtain dissociation between mechanical and electrical events, i e , electrical activity is obtainable when there is no longer a recorded mechanical response

The transmembrane potential and mechanical response of cardiac muscle are very sensitive to certain inorganic ions Changes in concentration of potassium, calcium, and to a lesser extent sodium, have a profound effect on the excitability and contractile response of the mammalian myocardium Magnesium has little action in the presence of normal calcium concentration, but when the latter concentration is decreased both magnesium and strontium have a marked effect on membrane potential

Transmembrane Potentials of Specialized Tissues. Of the three specialized tissues in mammalian heart, i e , the sinatrial node, the atrioventricular (AV) node, and the ventricular conducting system (Purkinje fibers), all have been studied by microelectrode techniques Figure 3 51*C* illustrates the electrical peculiarities of Purkinje fibers in the dog heart In contrast to the constant diastolic membrane potential in atrial and ventricular muscle, there is in the ventricular conducting system a slow depolarization throughout diastole when local electrical activity is induced by a distant pacemaker When pacemaker activity originates in the local area there is, in addition, a slow upward curvature in late diastole which continues until a level of instability is reached and a more rapid depolarization, the action potential, spontaneously begins (Fig 3 51*D*) This peculiarity was first described by Bozler (1943) Pacemaker activity in the sinoatrial and AV junctional tissue, is similar to the above when they act as pacemakers

As to be expected, the electrical properties of the pacemaker region change with factors that alter heart rate Vagus stimulation, acetylcholine, and hypothermia lower heart rate by decreasing the rate of diastolic depolarization and thus lengthening the time taken to reach a critical level at which an action potential is initiated, whereas sympathetic stimulation and epinephrine injection, by inducing a faster rate of

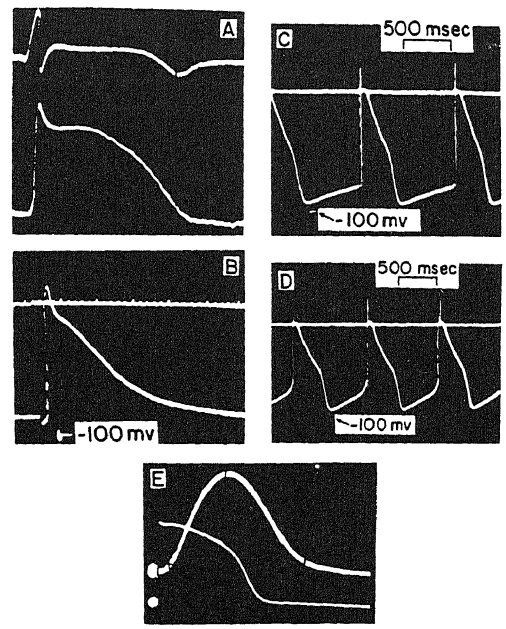


Fig 3 51 *A* Transmembrane action potential of a single ventricular fiber (lower curve) and simultaneous ventriculogram of the intact heart in situ The upstroke of the action potential coincides with the R wave of the electrogram the plateau with the R-T segment and the phase or repolarization with the T wave (After Brooks and associates) *B* Normal transmembrane potential recorded from a single fiber of the intact auricle (After Brooks and associates) *C* Transmembrane action potential of a single Purkinje fiber in a nonpacemaker activity Note slow loss of resting potential during diastole (After Brooks and associates) *D* Spontaneous activity recorded from Purkinje fiber in a pacemaker area Note slow depolarization during diastole and then abrupt onset of spontaneous depolarization (After Brooks and associates) *E* Temporal relation between action potential and mechanical contraction recorded from a papillary muscle (After Dudel and Trautwein 1954)

depolarization with an earlier firing, accelerate heart rate

Transmembrane action potential (TMP) curve during activation from various types of cells are diagrammed in Figure 3 52 Five phases of the action potential are recognized phase 0, an initial rapid depolarization or spike, which overshoots the zero potential by 20 to 30 mv , phase 1, an initial rapid repolarization which returns the TMP toward zero, phase 2, a slow phase of repolarization or plateau, phase 3, a final rapid repolarization which returns the membrane potential to its resting level, phase 4, the resting or diastolic TMP In pacemaker cells the TMP during phase 4 gradually decreases (diastolic depolarization) to a given potential (threshold) at which "all-or-none" depolarization occurs (phase 0)

Threshold Potential (TP). The critical level of cell membrane depolarization (about 55 to 60 mv) at which a propagated impulse is initiated (Fig 3 53) is the *threshold potential* Once this level of depolarization is reached, no greater stimulus is required for the upstroke of the action potential to become self-sustaining The *threshold stimulus* is defined as the magnitude of stimulus required to bring the resting membrane potential to the threshold level Subthreshold stimuli produce only local nonpropagated responses (Fig 3 53) As the resting TMP decreases toward TP excitability increases, i e , less TMP decrease is required to reach TP If the resting TMP becomes sufficiently less than the threshold potential, the cell is not excitable (i e , it cannot initiate a propagated impulse)

The form, duration, resting potential level, and amplitude of action potential are characteristic for the different myocardial cell types (Fig 3 52 *B*, *C*) These differ among mammalian species Those characteristic for man, dog, and rabbit will be described. The SA node pacemaker cells have a relatively low resting potential (Fig 3 52 *B*, *C*, and 3 54), slow phase 0, indistinct phase 1 and phase 2, distinct phase 3 and phase 4 slowly depolarizes Atrial muscle cells lack a distinct phase 2 (plateau) and phase 3 is prolonged Atrioventricular nodal cells may exhibit hyperpolarization at the end of phase 3 and phase 1 is not distinct Purkinje fibers have a distinct phase 1 and phase 2 is relatively long, phase 4 may exhibit spontaneous depolarization.

Most atrial and ventricular cells maintain a

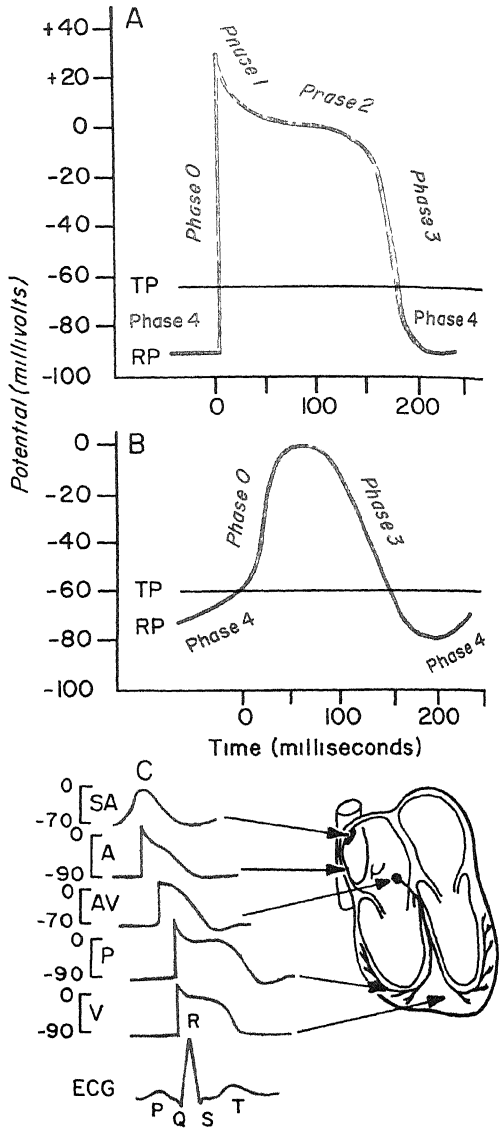


Fig 3 52 Transmembrane potential from a ventricular working or nonpacemaker cell (*A*) and a sinus node pacemaker cell (*B*) with phases labeled *C* illustrates the configurations of TMP from various types of cells from top to bottom sinus node cell, atrial muscle cell, atrioventricular nodal cell, Purkinje cell, ventricular contractile cell (Redrawn from Marriott and Myerburg, 1970)

stable resting membrane potential throughout the phase 4 and are termed *nonpacemaker cells*. Normally, they do not initiate impulses spontaneously

Pacemaker cells are located in the sinoatrial node, in tissues around but not in the atrioventricular node, Purkinje fibers, and atrium (la-

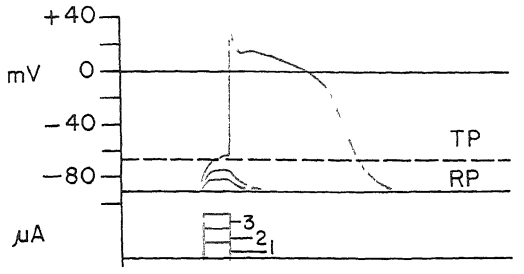


Fig 3 53 Changes in transmembrane potential produced by subthreshold (1 and 2) and threshold (3) stimuli *Upper* TMP *Lower* stimulus current With stimulus currents 1 and 2 graded local depolarizations occur but not a propagated response With current 3 the local response reaches threshold and a propagated TMP is initiated (Redrawn from Hoffman 1969)

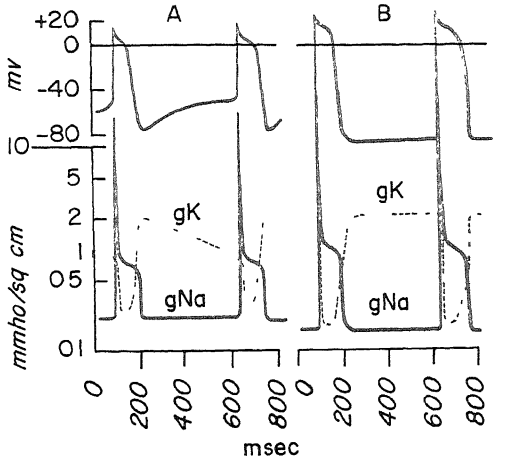


Fig 3 54 Course of estimated sodium and potassium conductance (g_{Na} g_K) curves related to the transmembrane action potentials of a sinus node pacemaker (*A*) and ventricular muscle (*B*) cell from isolated rabbit heart (Redrawn from Trautwein 1963)

tent atrial pacemakers, some of which, at least, have morphological characteristics of Purkinje cells) These cells have the capacity for progressive diastolic depolarization (phase 4)

MECHANISMS OF TRANSMEMBRANE POTENTIALS

Experimental observations and reasonable assumptions indicate that the ionic theory of Hodgkin and Huxley (1952) for nerve membrane potentials is valid for cardiac cells Myocardial cell membrane permeability and conductance to Na^+, K^+, and Cl^- account for transmembrane potentials

Membrane ionic *permeability* (P_i) is a measure of the ease with which an ion penetrates the membrane when driven by a concentration gradient across the membrane Membrane ionic *conductance* (g_i) is a measure of the ease of penetration of an ion when driven by an electrical force Ion permeability and conductance are related quantities, but not equivalent Their relationship is complicated, depending on such things as the membrane voltage and ionic concentrations on either side of it Ion conductance is the reciprocal of ion electrical resistance in a membrane ($g_i = 1/R$) and is expressed in mhos (ohms spelled backward) It reflects membrane ion permeability

Changes in P_i and g_i determine the membrane's potential (E_m) and flux or current (I_m)

The approximate concentration ratios intracellular extracellular for sodium, potassium, and chloride ions in heart muscle cells are

$$Na^+,\ 1\ 10,\ K^+\ 40\ 1,\ \text{and}\ Cl^-,\ 1\ 30$$

From the Nernst equation (q v) the equilibrium potential for a given ion (E_{ion}) can be calculated if the ionic concentrations inside (i) and outside (o) a cell membrane are known At 37° C

$$E_{(ion)} = 61\ 5 \log (ion)o/(ion)i$$

From typical values of ion concentrations for heart muscle, the estimated equilibrium potentials for Na^+ and K^+ are

$$E_{Na} = +70\ \text{mv}$$

$$E_K = -102\ \text{mv}$$

In Table 3 10 it is seen that the resting membrane potential for various myocardial cells approaches the equilibrium potential of K^+ Potassium ions have been identified as the major ion species responsible for maintaining resting potential That membrane potential (E_m) is not identical with E_K results from the influence of other ions, in particular Na^+ The degree of influence these ions have on resting potentials is determined by their conductance across the membrane To explain why E_m does not equal E_K consider the equation

$$E_m = 61\ 5 \log \frac{[K^+]_o + \alpha[Na^+]_o}{[K^+]_i + \alpha[Na^+]_i}$$

α is a constant expressing the ratio of Na^+

Table 3 10
Cardiac resting and action potentials

Site	Species	Fiber Diameter	MRP[a]	AP[a]	D_{AP}[a b]
		μ	*mv*	mv	*m /sec*
Sinus node	Frog	5	40	53	200
	Rabbit	—	60	66	200
AV node	Dog	—	53	58	350
Atrium	Frog	5	70	90	400
	Rabbit	10	78	92	150
	Dog	10	85	105	200
	Man	10	70	75	—
Ventricle	Frog	10	80	95	600
	Dog	16	80	100	400
	Man	16	87	115	200
Purkinje fiber	Dog	30	90	120	300
	Sheep	75	94	130	400
Skeletal muscle[c]	Frog	130	88	120	1 5
Unmyelinated nerve fiber[c]	Squid	400	65	85	0 4

From Hecht, 1968

[a] MRP, membrane resting potential, AP, action potential, D_{AP}, duration of action potential

[b] The values for D_{AP} depend on temperature and heart rate—some average values are given which show the difference in fiber size and in duration of the action potential in comparison to cardiac tissue

[c] Values for skeletal muscle and nerve fiber are given for comparison note particularly the differences in fiber size and in duration of the action potential in comparison to cardiac tissue

permeability to that of K^+ (P_{Na}/P_K) A small inward flux of Na^+ during rest, would oppose the outward flux of K^+ and reduce E_m, explaining why E_m is less than E_K Since α is thought to be about 0 01 for muscle and nerve, $\alpha[Na^+]_i$ is small compared to $[K^+]_i$ In the physiological range (low $[K^+]_o$ and high $[Na^+]_o$), $\alpha[Na]_o$ becomes a significant factor of the sum $[K^+]_o + [Na^+]_o$ Pacemaker tissues have a lower resting transmembrane potential than working myocardial cells, which could be accounted for by assuming a larger resting Na^+ inward current Cl^- inward flux has the same effect as K^+ outward flux and plays a role with K^+ during repolarization

The distribution of Na^+ and K^+ across the cell membrane is maintained by a metabolically active Na/K membrane pump which derives its energy from organic phosphate metabolism Adenosine triphosphatase (ATPase) located in the cell surface is activated by external K^+ and internal Na^+, i e , the membrane is asymmetrical with respect to Na^+ and K^+ Adenosine triphosphate (ATP) is the immediate source of energy for the pump, which is inactivated by metabolic inhibitors (dinitrophenol, azide, cyanide) and inhibited (although low concentrations can be shown to stimulate) by cardioactive glycosides (ouabain, digoxin) The latter show asymmetry of action too, inhibiting K^+ transport from the outside of the fiber This ATPase is Mg^{++} dependent and inhibited by Ca^{++} at the inner membrane surface

Resting Potential. The maintenance of resting potential is a result of the large ionic concentration gradients across the membrane The voltage and polarity of the resting potential are a consequence of the higher conductance of the resting membrane for K^+ than for Na^+ The resultant membrane potential, E_m, represents the sum of the equilibrium potentials for Na^+ and K^+, $E_{Na} + E_K$ Chloride ions have no effect on resting potential, but contribute to repolarization phases of the action potential Since potassium permeability (P_K) and conductance (g_k) are relatively high in the resting state, the resting membrane potential largely is dependent on K^+ flux

Action Potential. The action potential at a given membrane site results from sequential

changes in membrane ion permeability primarily to Na^+ and K^+, and to some extent to Cl^- The conducted cardiac impulse can be viewed as a traveling wave of ion permeability and conductance changes which transiently reverse membrane polarity The conductance changes for Na^+ and K^+ are graphically represented in Figure 3 54 At the onset of the action potential, membrane P_{Na} increases abruptly Na^+ ions flow down their electrochemical gradient into the cell, producing an action potential spike (phase 0) of about 100 mv There is evidence that Ca^{++} ions may also contribute to this inward current of depolarization During phase 1 g_{Na} decreases from its initial high level and g_K decreases Phase 2 is characterized by low g_K and a moderately high g_{Na} Repolarization occurs when the conductances of K^+ and Na^+ return to their resting levels Cl^- ion inward conductance apparently contributes to phase 3 In nonpacemaker cells TMP remains constant at the resting level during phase 4 This stability is largely the result of a steady g_K

The inward movement of Na^+ and outward movement of K^+ are passive movements requiring no energy, the ions flowing down their respective electrochemical gradients At each activation, minute amounts of Na^+ and K^+ are transferred across the cell membrane The metabolically active Na/K pump mechanism is located in the cell membrane It consumes energy from ATP split by Na/K activated ATPase This process returns K^+ to the cell and extrudes Na^+, against their respective concentration gradients

Pacemaker Potential. Pacemaker cells depolarize progressively during phase 4 (Fig 3 54) If not excited previously by a conducted impulse, they will discharge spontaneously when the threshold potential is reached The chief cause of diastolic depolarization is a time-dependent fall in g_K, i e , g_K decreases with elapsed time K^+ conductance is also decreased somewhat as the membrane potential falls during phase 4 (voltage dependence) This fall in g_K permits the low g_{Na} to become increasingly effective As g_K decreases during diastole, outward movement of Cl^- contributes to diastolic depolarization When the threshold potential is reached, g_{Na} increases suddenly and markedly, producing the spike potential (phase 0)

Refractory States Under physiological conditions, cardiac cells are refractory to further stimulation during most of the action potential period Excitability returns during the repolarization phase and goes through a period of supernormality Various phases of the refractory state are recognized (Fig 3 55)

The *absolute refractory period* (ARP) extends from the onset of the action potential until repolarization has progressed to a TMP of about 55 mv The cell cannot be excited during this period As repolarization continues there is a variable decrease in the stimulus required (threshold) and action potentials vary in rate of rise (phase 0) and amplitude (Fig 3 55) The *relatively refractory period* (RRP) is the interval during which action potential responses with slow phase 0 and submaximal amplitude occur Before a propagated response can occur, there is a period during which a stimulus may produce graded responses in cells which are not conducted The *effective refractory period* (ERP) is this period together with ARP The *total refractory period* (TRP) includes ARP,

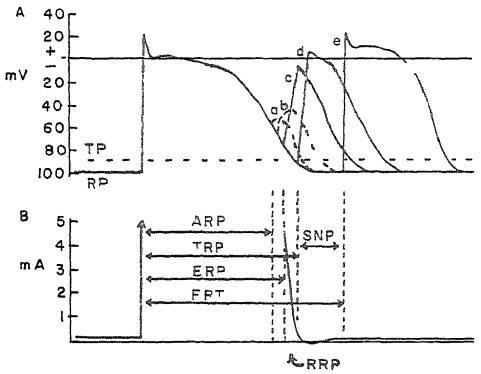


Fig 3 55 *A*, normal transmembrane action potential and responses to series of stimuli applied during and after end of repolarization ***B***, approximate durations of absolute refractory period (ARP) total refractory period (TRP) effective refractory period (ERP) full recovery time (FRT) supernormal period (SNP) and relative refractory period (RRP) In *A* responses shown as dotted lines (*a* and *b*) are graded responses which do not propagate Response (*c*) is earliest propagated response and defines end of effective refractory period Response (*d*) is elicited at the time when transmembrane potential is close to level of threshold potential (TP) and occurs during supernormal period Response (*e*) is elicited after end of repolarization and is normal in terms of rising velocity and amplitude, defining end of full recovery time Changes in threshold shown in ***B*** are related to an arbitrary scale of current strength Curve shows onset of inexcitability coincident with phase 0 of transmembrane action potential gradual decrease in threshold during phase 3 of repolarization and restoration of full excitability after end of phase 3 of transmembrane action potential (Redrawn from Hoffman 1969)

ERP, and RRP The *supernormal period* (SNP) is an interval beginning at about the time the action potential reaches the resting threshold level During this period the stimulus required for response is less than for a resting cell, but the amplitude and rate of rise of the action potential are submaximal *Full recovery time* (FRT) includes all of the foregoing periods up to the time when the resting potential is stable and the rise time, action potential amplitude and duration of the response and excitability are fully restored The term *partial recovery period* is sometimes employed for cardiac tissue or the entire heart to designate the interval during which some cells have recovered fully while others remain in some stage of refractoriness

Conduction Velocity. The rapidity with which the impulse is conducted in a given fiber is a function of the rate of rise and amplitude of the action potential and the fiber diameter The lower the resting potential, the slower is the speed of conduction The larger the fiber diameter, the faster is conduction velocity Sinoatrial and atrioventricular nodal fibers have lower resting potentials than atrial and ventricular cells and are smaller These fibers conduct more slowly

Decremental conduction is said to occur in sinoatrial and atrioventricular nodal tissues, i e , the action potential progressively decreases in rise time, velocity and amplitude, until propagation slows and finally ceases Under certain circumstances (e g , vagal stimulation) decremental conduction may increase in the AV node to the point that the impulse is blocked This may be a mechanism responsible for certain types of SA and AV block and so-called concealed conduction in AV junctional tissues (see under Arrhythmias and Conduction Disturbances)

AUTOMATICITY AND EXCITABILITY (IRRITABILITY)

The term *automaticity* designates the ability of cardiac cells to generate a propagated impulse spontaneously (assume pacemaker function) Normally, as stated earlier, this property is thought to be limited to specialized cells in the SA node, around but not in the AV node, in the atrium, and in the Purkinje system Electrophysiologically, these cells exhibit diastolic depolarization The rate at which diastolic depolarization occurs determines the frequency of spontaneous impulse formation, i e , the steeper the slope of phase 4, the shorter the time interval required for the TMP to reach the threshold potential required for spontaneous depolarization of the membrane When the firing frequency of a cell is increased, its automaticity is said to be increased and vice versa Certain abnormal mechanisms can cause automaticity For example, the drug aconitine causes oscillating afterpotentials to follow phase 3 of the action potential such an oscillatory depolarization can reach TP and initiate an action potential It is speculated that delayed repolarization of a group of cells may set up an excitatory current flow with already repolarized cells, sufficient to cause re-excitation of adjacent cells

In contrast to automaticity, the term *excitability* (or *irritability*) refers to sensitivity to stimulation The strength of stimulus required to bring the resting potential to the threshold for a propagated impulse, is determined by the difference between the existing membrane potential and its threshold potential (TP) If the resting TMP is greater than (i e , more negative) but close to TP, the stimulus needed to initiate an action potential is less than when the resting potential is much more negative than TP On the other hand, cells are refractory to stimulation when TMP is sufficiently less negative than TP (Fig 3 55)

Conduction System

The specialized pacemaking and conducting tissue of the heart consists of the sinoatrial node (SAN), interatrial and internodal conduction pathways, the atrioventricular node (AVN), the bundle of His, the right bundle branch (RBB) and left bundle branch (LBB) of the bundle of His, and the peripheral Purkinje network Ordinarily, pacemaker cells located within the SAN serve to initiate the heart beat From here, the impulse enters the atrial myocardium and follows preferential conduction pathways which conduct the impulse relatively rapidly to the atrioventricular node (internodal tracts) and to the left atrium (Bachmann's bundle) Normally, impulses which reach the ventricle must traverse the atrioventricular node, a region of relatively slow conduction Emerging from the AVN, the impulse reaches the rapidly conducting Purkinje fibers of the bundle of His and travels along the bundle branches to the subendocardial Purkinje fiber

network of the ventricles These structures serve to originate the cardiac impulse for each heartbeat and distribute it throughout the heart They provide for sequential contraction of the atria and ventricles and bring the entire myocardial mass of both ventricles rapidly into coordinated contraction Data on fiber diameters and conduction velocity in the human heart are given in Table 3 11 and the routes of conduction are presented schematically in Figure 3 56

Table 3 11
Fiber diameter and conduction velocity in human hearts

Site	Fiber Diameter	Conduction Velocity
	μ	*m /sec*
Sinus node	2-7	—
Atrium	3-17	0 8-1 0
AV fibers (about 5 mm long)	3-11	0 05
His bundle branch	9-18	2 0
Ventricular myocardium	10-12	
Subendocardium		1
Subepicardium		0 4-1

From Bauereisen, 1970

THE SINOATRIAL NODE (SAN)

The SAN is imbedded in the right atrial wall at the junction of the superior vena cava and the right atrium This junction appears on the epicardial surface as the *sulcus terminalis*, on the endocardial surface as the *crista terminalis* The SAN is ellipsoid in shape containing rather long thin fibers in which myofibrils are sparse An abundant collagen network interlaces with bundles of nodal fibers James (1970b) distinguishes two types of cells, centrally located P *cells* which may be the normal pacemaker cells and more peripherally located *transitional cells* which serve to connect P cells with adjacent cells of the atrial myocardium and the internodal and interatrial pathways (see p 3-28)

INTERNODAL AND INTERATRIAL PATHWAYS

These are diagrammed in Figure 3 56 Three pathways connecting the SAN to the AVN and one connecting the SAN to the left atrium have been described The anterior internodal tract travels from the left tip of the sinus node a short distance and divides into two branches one branch (Bachmann's bundle) travels along the dorsal aspect of the interatrial band and ramifies over the left atrium The other branch curves downward below the interatrial septum to the region of the AVN, here joining fibers from the other internodal tracts The middle internodal tract (first described by Wenckebach) emerges from the posterodorsal edge of the sinus node, passes behind the superior vena cava to the top of the interatrial septum Here, a few branches pass to the left atrium while the major bundle continues over the interatrial septum to the top of the AVN The posterior internodal tract (Thorel's tract) extends from the posterior end of the sinus node along the *crista terminalis*, through the eustachian

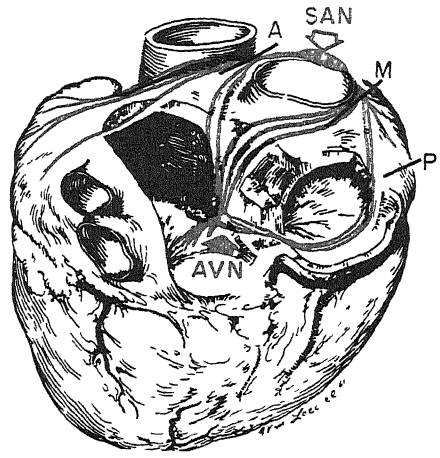


Fig 3 56 Diagram of the internodal pathways of the human heart ***A***, anterior internodal pathway which sends fibers to the AVN and interatrial fibers to the left atrium (Bachmann s bundle) ***M*** middle internodal tract (Wenckebach s tract) ***P***, posterior internodal tract (Thorel s tract) (James 1970a)

ridge to the interatrial septum to join the AVN at its right superior margin

These internodal pathways constitute preferential routes of conduction from the SAN to the AVN and from the SAN to the left atrium Conduction velocity is more rapid in these pathways than in ordinary working atrial myocardium but not as rapid as in the His bundle or its branches The pathways consist of a continuum of myocardial cells made up of a mixture of Purkinje cells and cells indistinguishable from ordinary working myocardium On anatomical grounds, the most direct route for spread of excitation from the SAN to the AVN and left atrium would be through the anterior internodal tract, one branch of which sends

specialized fibers (Bachmann's bundle) into left atrium and the other branch into the AVN Conduction is most rapid in this tract in dogs (Holsinger et al , 1968) Atrial depolarization is accomplished both by radial spread to working myocardial cells from the SAN as well as more rapid transmission over these preferential pathways directed toward the AVN and left atrium The internodal tracts are not ensheathed nor otherwise separated from adjacent myocardial tissue These special pathways provide for rapid spread of excitation throughout both atria and determine its temporal sequence in different parts of the atria There is evidence that specialized atrial cells exhibit elevated resistance to the depressant effects of hyperkalemia and can continue to conduct impulses from the SAN to the AVN when the body of the atrial musculature is inactivated and the P wave of the body surface electrocardiogram absent (Vassalle and Hoffman, 1965, Wagner et al , 1966)

ATRIOVENTRICULAR NODE

The atrioventricular node is located beneath the right atrial endocardium at the base of the interatrial septum just interior to the ostium of the coronary sinus (Fig 3 57) On the atrial side, it receives branches from the internodal tracts as well as from the ordinary atrial musculature and is continuous with the bundle of His on the ventricular side Its fibers, which are smaller than those of atrial muscle, form a pattern of branching and interlacing strands in a connective tissue framework P cells are present within the node especially in its deeper portions, but they are less frequent then in the sinus node The slender transitional cells are the most frequent cell type

CONDUCTION TO VENTRICLE

The Bundle of His At the region of transition between the AVN and bundle of His, the cells gather into more regular fascicles and take on the characteristics of typical Purkinje fibers These fibers of the His bundle then pass from the right atrial endocardial location of the AVN medially and inferiorly along the lower margin of the membranous interventricular septum (Fig 3 57) The His bundle gives rise to a single right bundle branch and multiple left branches, the latter forming a sheet of fibers traveling down the left septal endocardium

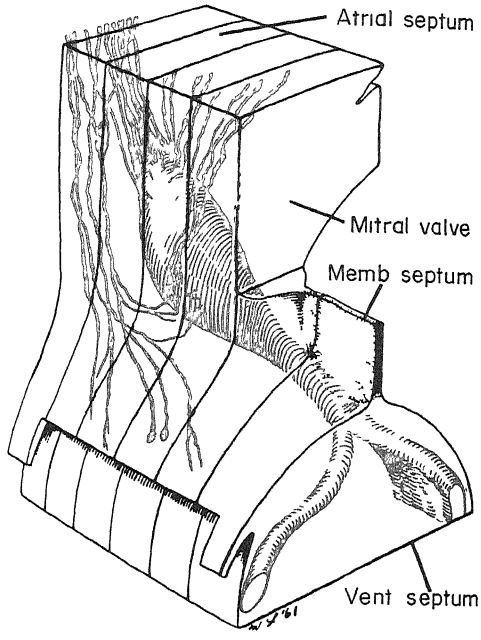


Fig 3 57 Schema of the AV node and His bundle Atrial Septum interatrial septum Vent Septum interventricular septum Memb Septum membranous portion of the interventricular septum Fibers from internodal tracts are shown entering the AV node The bundle of His (not labeled) runs along the lower margin of the membranous interventricular septum It gives rise to a single right branch and then multiple left branches form a vertical sheet down the left septal subendocardium (James 1970a)

The Bundle Branches. The right bundle is given off first as the smaller of the two branches It passes down, subendocardially, the right side of the interventricular septum and, branching, links up with the subendocardial Purkinje network The first functional contact of the right bundle branch with the ventricular myocardium is near the base of the anterior papillary muscle

The left bundle branch continues from the common bundle of His as a broad fascicle which runs subendocardially down the left side of the interventricular septum, dividing into major anterior and posterior divisions before reaching the peripheral Purkinje network In most hearts the left bundle branch makes its first functional contact with the ventricular myocardium on the left endocardial surface of the interventricular septum some distance below the aortic valve

The paraspecific fibers of Mahaim are incon-

stant, but when present are seen as a small bundle of fibers leaving the left bundle branch to enter the upper dorsal parts of the interventricular septum

The Accesory Bundle of Kent Additional communications between atria and ventricle, most frequent in fetal and infantile hearts, may be identified in some adult hearts The most common is the bundle of Kent which extends from the right atrium to the upper parts of the interventricular septum Fascicles of atrial myocardium may also be seen extending through the atrioventricular ring into the base of the atrioventricular valves Purkinje cells are not found in the accessory bundles They are considered by some to serve as atrioventricular conduction routes, by-passing the AVN, which, when active, account for accelerated atrioventricular conduction (Wolff-Parkinson-White syndrome)

The Subendocardial Purkinje Network. The Purkinje network is largely limited to the subendocardium in man These specialized conducting cells of Purkinje penetrate for a short distance into the ventricular myocardium where they merge insensibly with working myocardial cells These fibers are also found in the free-running false tendons in the right and left ventricles

IMPULSE INITIATION

Keith and Flack, 1907, suggested that the SA node embedded in the right atrial wall near the entrance of the superior vena cava is the focus in the mammalian heart from which each heart beat normally starts Evidence of such origin is that *1)* its excision or inactivation causes cessation or slowing of the atrial or ventricular contractions, *2)* temperature change of the node but not of the surrounding tissue changes heart rate, *3)* the node becomes electrically active before any other region of the heart

Although spontaneous and repetitive depolarization arises normally only in this region, there are other areas in which electrical activity can arise under some circumstances and initiate cardiac contraction When the SA node is destroyed or inactivated, the AV junctional tissues often become the site of impulse formation and develop spontaneous depolarization Another area which develops pacemaking activity when the SAN is inactive is in the region of the coronary sinus In turn, in the presence of inactivation of the AV junctional tissue, the ventricles develop a slow rhythm Since both ventricles receive the excitation through natural pathways and contract in a coordinated way, the rhythmical center is the AV bundle If both bundle branches are cut, impulse initiation arises in the ventricular muscle The natural rate of automaticity of these areas is of a descending order of magnitude, the SA node being 70 to 80/min , atria 60/min , AV junctional tissues 40 to 60/min , and ventricles 20 to 40/min It is obvious that normally, the atria, AV junctional tissue, and ventricular tissue cannot set the heart rate because the SA node repolarizes and automatically discharges more rapidly than any of these tissues Hence, an impulse from the SA node will always depolarize these tissues before any could initiate a rhythm On occasion, however, an ectopic focus in the atrium or ventricular muscle can assume a rhythm faster than that of the SA node, and hence, it becomes the pacemaker

SPREAD OF EXCITATION IN HUMAN HEART

The cardiac impulse normally is initiated in sinoatrial node pacemaker cells (Fig 3 52) Thus, the sinus node sets the pace and depolarizes other potential pacemaker cells before they can discharge automatically If the frequency of sinus node cells decreases greatly or they fail to discharge automatically, other slower pacemakers will begin to initiate the heart beat Ectopic foci in the atria or ventricles can develop a faster rhythm than the sinus node cells under certain conditions These sites then determine cardiac rate and rhythm from an abnormal location

Although preferential pathways of conduction have been identified in human and dog hearts, as described in the foregoing, in isolated human hearts spread of excitation proceeded more or less radially from the region of the SA node at a rate of about 1 m /sec (Durrer et al , 1970) Conduction through the atrioventricular node is slow and decremental, requiring 80 to 120 msec Here conduction velocity may be increased by sympathetic stimulation, ordinarily traveling at a velocity of 0 1 to 0 2 m /sec The Purkinje fibers of the bundle of His and bundle branches conduct rapidly at rates of about 2 m /sec , distributing the impulses fairly synchronously to the subendocardium, where conduction velocity is about 1 m /sec Excitation is then conducted from endocardium to epicar-

dium Conduction velocities in the human ventricular myocardium of about 0 35 to 0 40 m /sec have been found (Durrer et al , 1970)

During the first 5 msec of ventricular excitation depolarization of the left central interventricular septal region and anterior and posterior paraseptal surfaces begins and these areas become confluent in 15 to 20 msec Subendocardial depolarization in the right ventricle begins near the insertion of the anterior papillary muscle 5 to 10 msec after the onset of ventricular activation Septal activation starts in the middle third of the left side of the interventricular septum and at its lower third near the septal posterior wall junction After 30 msec , a nearly completely confluent front is moving through the left ventricular wall, except at the posterobasal wall, and subepicardial excitation has appeared on the anterior surface of the heart near the atrioventricular sulcus and at anterior and posterior paraseptal areas halfway between apex and base The posterobasal area is the latest to be activated Right ventricular activation, starting at 5 to 10 msec , rapidly invades the septum and adjoining right free walls reaching the subepicardium in 20 msec near the area pretrabecularis Ultimately, the pulmonary conus and posterobasal areas are activated (at 60 to 70 msec) Thus, the general sequence of ventricular activation is *1*) septal excitation proceeding from left to right, *2*) excitation in the upper part of the anterior wall of the left ventricle, proceeding through the free wall, *3*) excitation of the left ventricular posterior wall proceeding in the opposite direction, *4*) excitation traveling through the right ventricular free wall and moving from right to left through the midanterior part of the septum, *5*) completion of activation at the pulmonary conal and posterobasal areas (Durrer et al , 1970)

ELECTROCARDIOGRAPHY

Electrocardiography may be defined as that branch of physiology which is concerned with the recording and analysis of the electrical activity of the heart The electrocardiograph is an instrument which receives the electrical impulses as they vary during the heart cycle and transforms them into a graphic record called the *electrocardiogram*, when recorded from the surface of the body Records obtained directly from cardiac muscle are called *electrograms*

Instrumentation

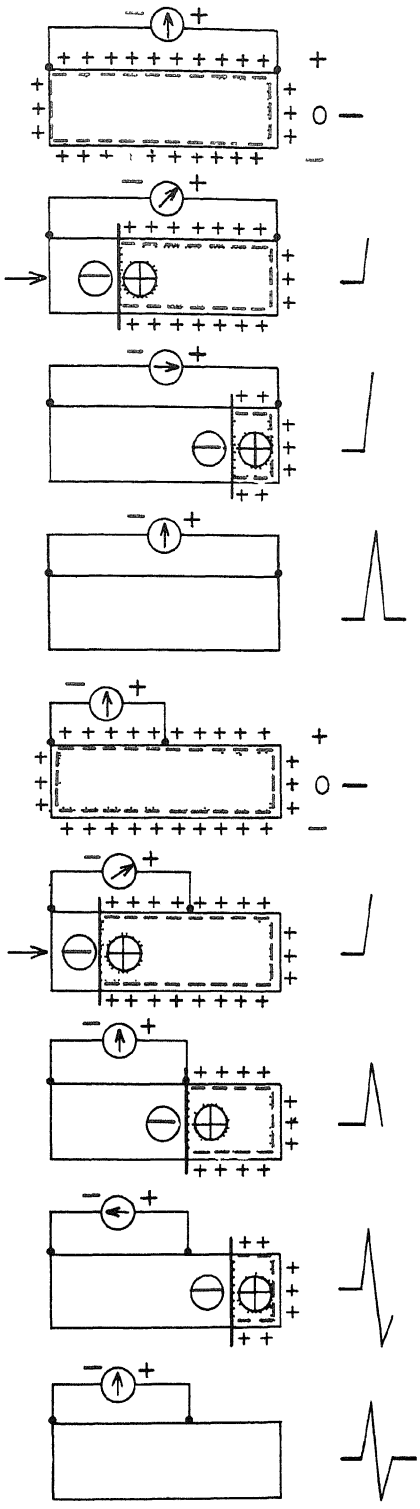
The first electrocardiogram was obtained in 1887 by Waller who employed a capillary electrometer Einthoven became the real father of electrocardiography when he described the string galvanometer in 1903 This instrument became the standard apparatus for the recording of electrocardiograms until it was largely supplanted by the vacuum tube, amplifier-driven oscillograph in the 1930's This type of instrument was later modified by increasing the amplification sufficiently to drive a mechanical lever arm so that directly written records might be obtained either with an ink writer or a hot stylus applied to specially waxed paper Recently transistorized modifications of the vacuum tube type of oscillographic instrument have come on the market Common to all these instruments is the property of faithfully inscribing the variations of the cardiac voltages on an accurate time scale abscissa

The cathode ray oscilloscope has been employed for specialized research in electrocardiography and recently it has been adapted by means of the sweep circuit to give a constant visual display of the electrocardiographic pattern which can be monitored during surgery or other critical clinical conditions

Depolarization and Repolarization of Cell Aggregates

The electrocardiogram which is recorded from body surface electrodes is the result of a vast number of systematically propagated electrical events taking place in the individual muscle cell fibers of the heart We have seen that the individual myocardial fiber is depolarized and repolarized with large voltage fluctuations across the cell membrane, but with only small potential differences being recorded outside the cell How then does the transmembrane action potential relate to the electrocardiogram recorded from the surface of the heart or from the surface of the body? One simplified approach to understanding the complex phenomena involved will be here presented

It must be remembered first that the depolarization of a portion of a fiber initiates a flow of current from the polarized zone to the depolarized zone. This flow of current propagates the wave of depolarization until the entire tissue is depolarized Since the heart is functionally a syncytium of muscle fibers, if one portion of the



muscle becomes depolarized it will initiate a wave of activation that will eventually depolarize the entire heart That such is indeed the case is shown by the fact that extrasystoles may be produced by touching the epicardium of the exposed heart at surgery or by touching the endocardial surface with a cardiac catheter during diagnostic studies The localized point of depolarization initiated by the mechanical stimulus is propagated to the entire heart muscle

THE ISOLATED MUSCLE STRIP

Depolarization The relationship between the surface electrocardiogram and the spread of depolarization through the muscle is illuminated by experiments on isolated strips of heart muscle Figure 3 58 illustrates the sequence of events when a wave of activation sweeps over a strip of myocardium with electrodes applied directly to its surface On the resting polarized muscle there is no difference of potential since all parts of the surface are equally positive in charge Hence the galvanometer records only a straight line at zero potential When the strip is stimulated at the region of the negative electrode to the left of the drawing, this zone becomes depolarized and is electrically neutral The other end of the strip is now positive with respect to the depolarized zone, and the galvanometer records a positive deflection As the wave of activation approaches the positive electrode more and more positivity is recorded until the muscle strip is completely depolarized, whereupon there is no longer a potential difference, and the recording drops abruptly to the zero line

When the positive electrode is placed in the middle of the strip and the stimulus is applied to the region of the negative electrode, posi-

Fig 3 58 Depolarization of a strip of myocardium with electrodes in direct contact with tissue During the resting stage no difference of potential is recorded When stimulated at arrow depolarization begins and stimulated end becomes negative relative to opposite end Positivity is recorded from positive electrode until activation is completed whereupon no more difference in potential exists and recording drops to zero With positive electrode in middle of strip positivity is recorded until the activation wave arrives under the electrode Since the plane that divides a dipole is at zero potential the recording drops immediately to zero As soon as the dipole passes the electrode maximum negativity is recorded and it gradually lessens until the return to zero when activation is completed

tivity is recorded until the wave of activation passes beneath the positive electrode whereupon the recording drops abruptly to a negative potential which returns to the zero line when the strip becomes completely depolarized In a series of simple but beautifully conceived experiments, Craib (1927) was able to show that this phenomena could best be accounted for in terms of the migration of a dipole across the length of the muscle strip

A dipole may be conceived of as two points of opposite electrical charge separated by a very small distance If these points are connected by a wire it is understood by convention that current will flow along the wire from the positive pole to the negative pole (In point of fact the electrons actually flow from the negative to the positive pole) If the dipole is immersed in a volume conductor such as a pool of weak saline solution, then a three dimensional field will be set up with the lines of current flow distributed in a pattern as illustrated in Figure 3 59 Isopotential lines are arranged at right angles to the lines of current flow Thus, all points which lie on the plane perpendicular to the center of the axis connecting the two points of the dipole are at zero potential The human electrocardiogram may be analyzed in terms of the body surface potentials generated by a series of dipoles within the heart and distributed throughout the body which behaves as a volume conductor

This concept may be clarified by reference to an isolated heart muscle strip immersed in a volume conductor as shown in Figure 3 60 In the resting state the strip is polarized with innumerable dipoles distributed about the surface of the strip, positive charges on the outside and negative charges on the inside Since the positive charges facing the electrode are exactly balanced by the positive charges facing away from the electrode, there is no difference of potential, and no deflection is registered on the galvanometer or on the electrographic tracing

When a stimulus is applied to the end of the strip facing away from the electrode, depolarization begins at that point The dipoles at the point of stimulus are obliterated by a flow of current across the stimulated membrane, the electrical resistance of which drops precipitously The stimulated area is now electrically negative with reference to the remainder of the strip which is still positively charged Thus, the plane that divides the depolarized from the polarized portion may be considered as an interface separating the positive and negative

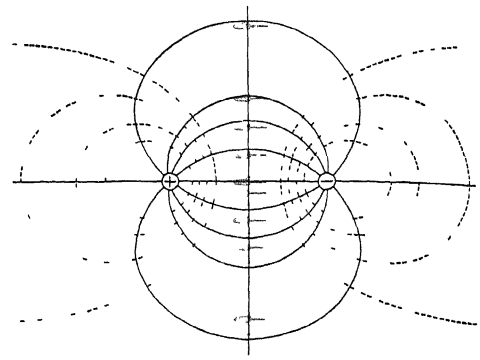


Fig 3 59 Electrical field of a dipole in a volume conductor Solid lines are current flow Dotted lines are isopotential lines

poles of a dipole, and the propagation of the wave of depolarization down the muscle strip may be thought of as the sweep of a dipole across the length of the muscle strip As the dipole approaches the electrode increasingly larger voltages are recorded as the electrode intersects isopotential lines of higher voltage Maximum potential is reached when the dipole arrives at the end of the strip nearest the electrode When the entire strip becomes depolarized, it suddenly becomes electrically inert and the galvanometer and electrogram immediately fall to the zero level

Figure 3 61 illustrates the electrograms which are inscribed when the positive electrode is placed at the stimulated end of the strip or at the midpoint

Repolarization. The process of repolarization is of relatively long duration in contrast to the rapid spike of depolarization, as demonstrated by the monophasic action potential recorded from single fibers with the micropuncture technique Since the resting transmembrane potential is fully restored after each beat, the total electrical charge of repolarization must be equal and opposite to the discharge of depolarization If it were presumed that repolarization followed the same pathway across the muscle strip as the wave of depolarization, then one would expect that the repolarization wave should be inscribed by the recording oscillograph as a slow deflection opposite in direction to the depolarization spike but encompassing the same area While this is indeed the case in the isolated muscle strip electrogram, it is not the case in the normal human electrocardiogram in which the depolarization and repolarization deflections usually have the same direc-

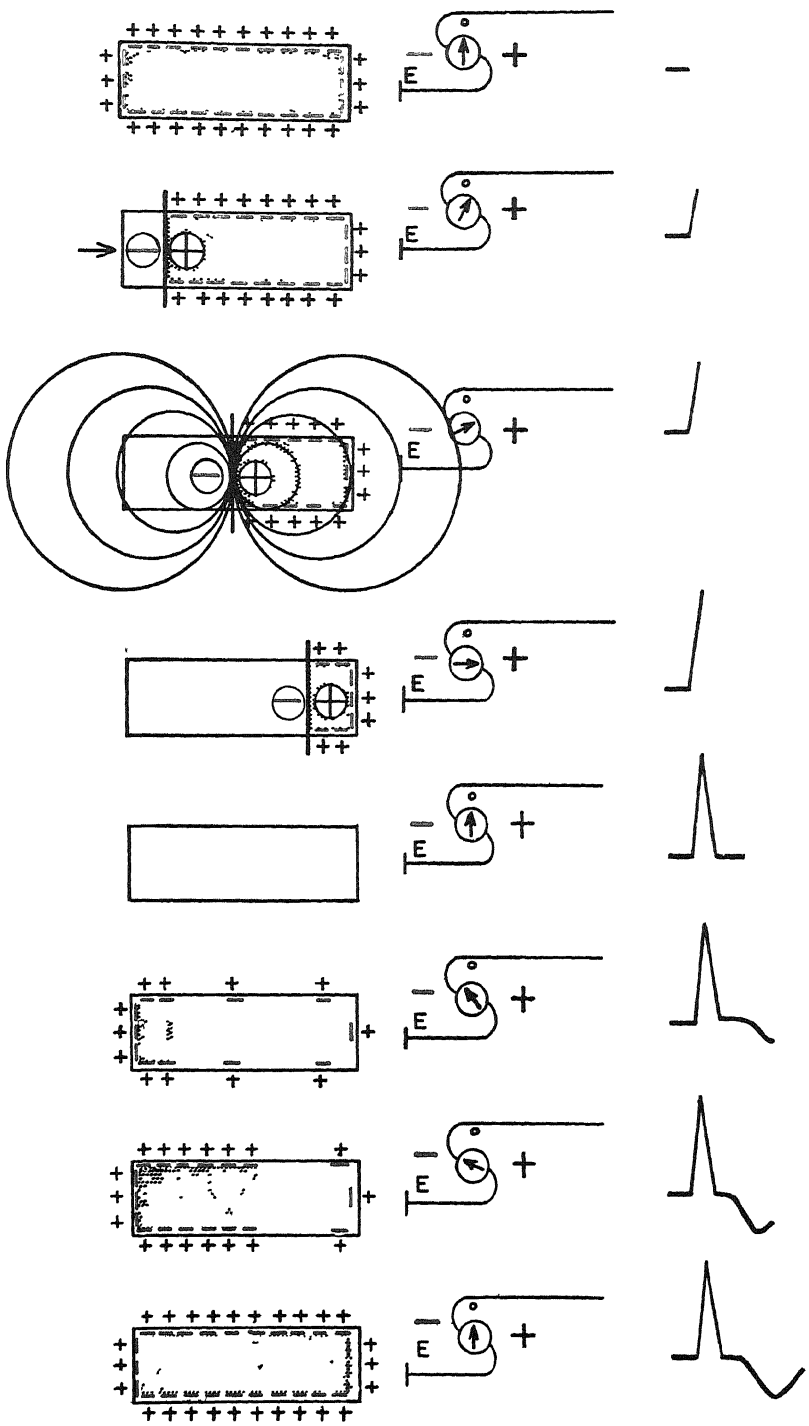


Fig 3 60 Depolarization and repolarization of a strip of myocardium immersed in a volume conductor Electrode E connected to positive pole of recording galvanometer which is paired with a distant electrode at near zero potential The advancing dipole at the border between activated and inactivated muscle produces an electrical field which moves toward the electrode The electrode is intersected by isopotential lines of increasing force until all difference of potential is extinguished by completion of depolarization Recovery occurs in all parts of the muscle nearly simultaneously but is completed at the stimulated end first thus producing relative negativity at the side facing the electrode and giving rise to a negative repolarization wave

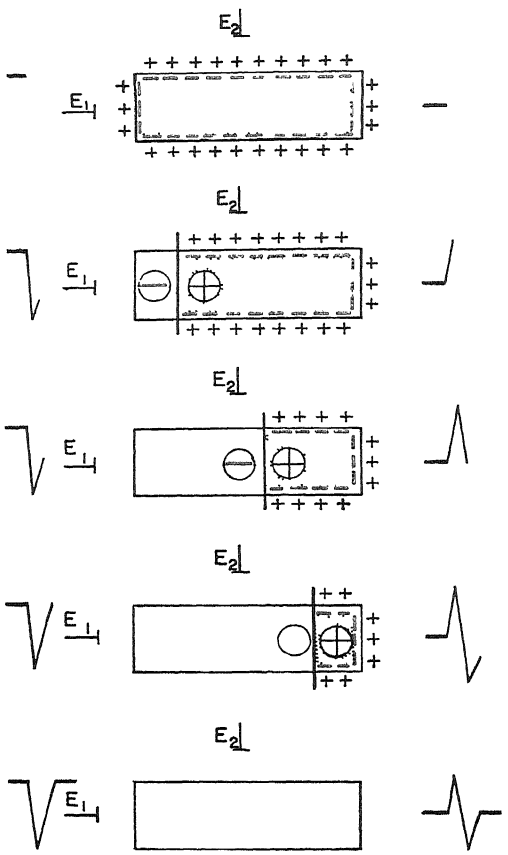


Fig 3 61 Form of complexes inscribed from electrodes placed at positions E_1 and E_2 during depolarization of a strip of myocardium suspended in a volume conductor Each electrode paired with a distant indifferent electrode Tracings on the left side are related to E_1 while tracings on the right relate to E_2 electrode

tion, and the areas of the two are rarely if ever the same The reasons for the paradoxical orientation of repolarization in the human electrocardiogram will be discussed in a subsequent section

Since repolarization is of long duration, different phases of it are proceeding simultaneously in various portions of the muscle strip and indeed throughout the heart (Fig 3 60) Repolarization, therefore, cannot be analyzed as a sequence of propagated electrical events and cannot be viewed as a wave front sweeping across the muscle Changes in temperature, chemical environment, and pharmacological agents may produce marked alterations of the repolarization wave without noticeably affecting the depolarization spike

Current of Injury. When a portion of myocardium becomes injured, whether it be mechanically, chemically, or thermally, the cell membranes in the injured area become more permeable to ions, and this leads to a lessening of the ionic concentration gradients of sodium and potassium across the cell membranes and thus to a reduction of the charge density in the resting polarized state This reduction of positivity in the injured area produces a gradient of charge from the uninjured to the injured zone A current flows from the area of high positivity to the area of low positivity, this current is called the current of injury It flows during the resting polarized state

Figure 3 62 shows how the injury current affects the depolarization and repolarization waves of the electrogram of the isolated muscle strip Prior to injury the polarized muscle strip gives rise to no potential difference and the electrogram tracing is at zero Upon injury of a portion of the strip near the electrode, the base line shifts in a negative direction since the charge density at the injured area is less than that over the remainder of the strip The injured area is thus electrically negative to the noninjured area When the strip is stimulated from the noninjured side the dipole of the depolarization wave sweeps toward the electrode producing a positive deflection The tracing shows an increasing positive spike until the entire strip has become depolarized, whereupon the electrogram deflection falls to the original zero potential, which is well above the negatively displaced base line of the depolarization spike Repolarization then begins at the zero potential line giving the appearance of elevation of the repolarization wave

If the electrode is placed on the opposite side of the strip, then the injury potential displaces the base line upward from the zero line The depolarization spike is negative, and its termination at zero falls below the point of origin This gives the appearance of depression of the repolarization wave Thus, the injury potential produces elevation of the origin of the repolarization wave when the electrode faces the injured zone and depression of the origin of the depolarization wave when the electrode faces the noninjured zone

Another form of injury potential may become apparent when the injured zone fails to completely depolarize As the dipole of the depolarization advances toward the electrode, increasing positivity is inscribed until the activation wave arrives at the border of the injured

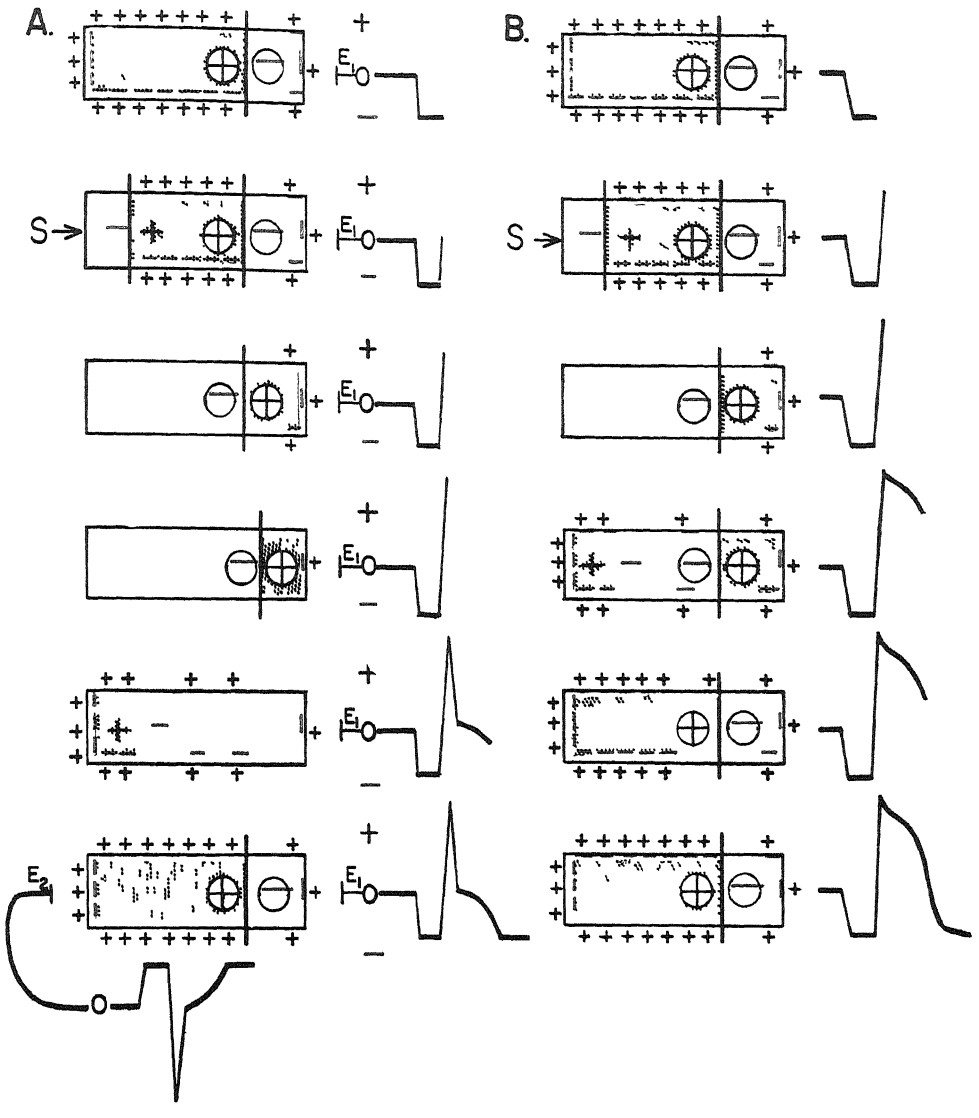


Fig 3 62 Sequence of electrical changes in myocardial strip which has been injured at end near electrode E_1 When portion of strip becomes injured there is leakage of ions and reduction of surface charge density making injured zone less positive than noninjured zone A dipole is thus produced at border of injured zone with negative pole facing the electrode during the resting phase The base line is therefore displaced downward to negative position Following stimulation the dipole of depolarization sweeps toward electrode producing positive force which neutralizes effect of injury dipole In *A*, the injured zone becomes depolarized and injury dipole is abolished Upon completion of depolarization recording falls to original zero position As repolarization is completed from left to right the injury dipole is re-established and tracing returns to negatively displaced base line giving appearance of elevation of onset of repolarization deflection If electrode is on opposite side of strip base line is displaced upward and onset of repolarization appears to be displaced downward *B* shows the sequence when injured zone fails to depolarize The impulse is blocked at the boundary of the injured zone which remains polarized and therefore during systole presents a dipole with its positive pole facing the electrode holding the recording in a positive position until repolarization returns the tracing to the negative base line This phenomenon is called the monophasic injury wave

zone where it is blocked At this point, the injured zone maintains it weakly polarized state, but the noninjured zone has become depolarized making it electrically negative to the injured zone The persistence of the dipole at the completion of activation maintains positivity at the electrode until the repolarization process in the uninjured zone slowly returns the tracing to

the isoelectric line The resulting complex representing a fusion of depolarization and repolarization waves has been called the monophasic injury wave (Fig 3 62*B*)

A third explanation for the injury potential is that the injured tissue repolarizes more rapidly than normal and therefore has a positive charge in relation to the surrounding normal tissue during an early part of the recovery period An electrode near the injured zone faces relative positivity during this part of the recovery period and thus records elevation of the origin of the repolarization wave

Mammalian Electrocardiogram

NOMENCLATURE

The mammalian electrocardiogram contains three major components during each cardiac cycle (Fig 3 63) According to the nomenclature devised by Einthoven, the component produced by atrial activation is called the P wave, the one produced by ventricular activation is the QRS complex, and the component produced by ventricular recovery is the T wave The P wave is separated from the QRS complex by a usually isoelectric interval called the PR segment Atrial repolarization begins after the first portion of the P wave and may produce a visible deflection of the PR segment called the Ta wave In man it is generally not identifiable but may be prominent in certain larger species such as the horse When present, it is usually in an opposite direction from the P wave and may extend beyond the QRS complex, deviating the initial portion of the ST segment (q v) This effect of a large Ta wave can be mistaken for abnormal ST segment deviation Individual deflections in the QRS complex are labeled as follows if the initial deflection is negative (downward from the base line) it is a Q wave, the first positive (upward) deflection is an R wave, and a negative deflection following the R is an S wave Subsequent positive and negative deflections in this complex are labeled R′, R″, etc , and S′, S″, etc , respectively The interval between QRS and T is the ST segment The junction between the end of the QRS complex and the ST segment is called the ST junction or J point A small afterpotential called the U wave is sometimes seen following the T. The duration of the components and intervals vary with species and age

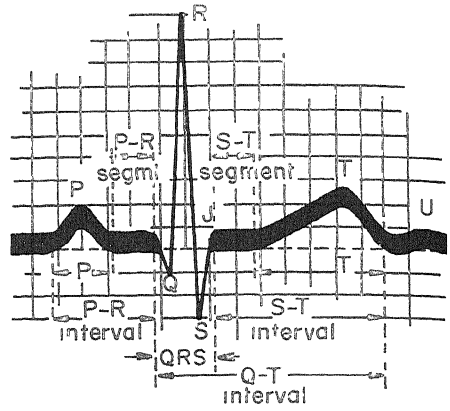


Fig 3 63 A single electrocardiographic complex giving nomenclature of the deflections and the intervals (Burch and Winsor 1966)

The intervals also vary inversely with heart rate In the normal adult human the P wave is 0 07 to 0 12 sec in duration, the PR interval measured from the beginning of the P wave to the beginning of the QRS complex is 0 12 to 0 20 sec , the QRS complex is 0 05 to 0 10 sec , and the QT interval measured from the beginning of the QRS complex to the end of the T wave is 0 26 to 0 45 sec at normal heart rates

DISTRIBUTION OF ELECTRICAL FORCES IN THE BODY

In accordance with dipole theory, the wave front of activation is considered to be a polarized surface with positive charages on its leading side and negative charges on the trailing side The axis joining each pair of positive and negative charges is perpendicular to the polarized surface, and the density of charges is considered to be a constant Atrial and ventricular activation are schematically described by the development of polarized surfaces which spread across the myocardium and are extinguished when they reach a margin of the myocardial syncytium or an area of activated muscle Laws governing the distribution of electricity in a volume conductor define the electrical field produced in the body by such polarized surfaces

When viewed from a sufficiently distant point P, a polarized surface can be represented by a single central resultant dipole with its axis perpendicular to a plane intercepting the boundary of the surface, and its charge proportional to the area of this plane (Fig 3 64*A*)

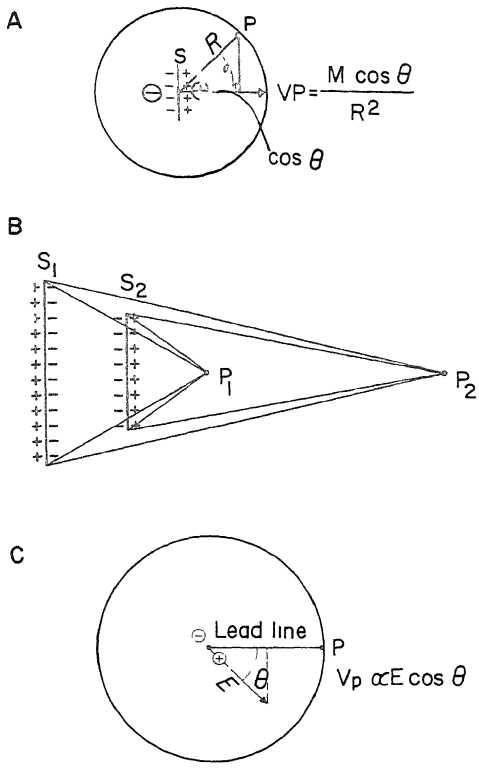


Fig 3 64 ***A*** Calculation of positive potential (V) at point P when polarized surface S is considered as a single dipole Arrow indicates axis of dipole If P is located on a unit sphere then R = 1 As P moves away from the dipole axis its potential diminishes as cos θ diminishes from 1 toward zero The numerical value of cos θ is given by the distance on the axis from the center of the circle to the perpendicular dropped from P M = dipole moment ***B,*** illustration of solid angle principle Two charged surfaces S_1 and S_2 have similar charge density but different areas and opposite orientation of charge At point P_1 the net potential is positive because S_2 subtends a larger solid angle than S_1 At point P_2 the net potential is negative because S_1 subtends the larger solid angle ***C,*** determination of voltage at point P when strength of dipole is represented by vector E Voltage is proportional to E cos θ and is found by drawing perpendicular from the lead line to the terminus of vector Voltage at P is determined by the distance measured along lead line from center of dipole to the perpendicular

The electrical potential (V_p) at a distant point in a large volume conductor containing a dipole with charges close together is proportional to M cos θ/R^2, where M is the electrical force of the dipole (determined by the quantity and distance between charges), R is the distance from P to the center of the dipole, and θ is the angle between the dipole axis and the line from P to the dipole center A positive potential is recorded at P when it faces the positive side of the polarized surface The direct relationship between V_p and cos θ indicates that potential at P is maximal when the dipole axis is parallel to R and zero when it is perpendicular to R V_p is also proportional to the solid angle at P which is subtended by the boundary of the polarized surface (Fig 3 64*B*) This solid angle is equal to the area cut on a unit sphere by lines drawn from its center at P to all points on the boundary of the polarized surface When two or more polarized surfaces subtend solid angles at P, the resultant V_p is proportional to the sum of the solid angles, with due regard for their sign

The electrical potential at any point in the body or on its surface can theoretically be predicted in terms of the solid angles subtended by polarized surfaces developed in the heart during activation, and this has been done in a qualitative fashion However, it is impossible to reverse the procedure and derive information about the size and orientation of the individual polarized surfaces from a knowledge of the resultant solid angle It has been more practical to analyze tracings recorded at a distance from the heart in terms of resultant vectors which indicate the magnitude and mean direction of electrical forces developed during the cardiac cycle without any implications about the polarized surfaces responsible for them When viewed from a distance, the total electrical forces of the heart at any instant can be considered as a single dipole that is described by a vector oriented in the direction of the dipole axis with magnitude proportional to the dipole moment This vector represents the sum of many individual vectors each of which describes a single dipole at a polarized surface

VECTOR ANALYSIS AND ELECTROCARDIOGRAPHIC LEADS

A vector quantity is represented geometrically by an arrow with its direction and length related to the direction and magnitude of the quantity it describes It can be shown that electrical potential (V_p) at a point P in a volume conductor is proportional to the geometric projection of an arrow representing a dipole vector onto a line (the lead axis or lead line) drawn from P to a point of zero potential at the center of the dipole V_p is proportional to E cos θ when E is the magnitude of the dipole vector and θ is the angle between the vector and the

lead line (Fig 3 64*C*) An electrocardiograph will record this potential if one electrode is placed at P and the second electrode is placed at a point of zero potential However, when two electrodes are placed on the body surface, the tracing recorded represents only the difference between potentials of the same order of magnitude at each electrode and does not give any specific information about the potentials at either point

Einthoven described a method for vector analysis of such tracings recorded from pairs of electrodes on the body surface at a distance from the heart He assumed as a first approximation that electrical forces of the heart originate in a relatively small area at the center of a homogeneous volume conductor, and that the regions where each arm and the legs join the trunk are points equidistant from each other The extremities behave as linear conductors connected to the trunk so that an electrode placed anywhere on an extremity is equivalent to one at its junction with the trunk Electrodes placed on each arm and a leg are then considered to be located at the apices of an equilateral triangle with the heart at its center, and differences in potential recorded between these points represent the projection of vector forces originating from a dipole at the center of the triangle onto lead lines drawn between its apices Potential differences between the extremities are customarily recorded in three standard limb leads obtained by connecting two of the limb electrodes at a time to the electrocardiographic instrument in the manner illustrated in Figure 3 65 The polarity of the electrode connections is such that a positive charge facing the left shoulder will produce a positive deflection (upward from the base line) in lead I, and a positive charge facing the foot will produce positive deflections in both leads II and III Einthoven deliberately arranged the polarity of lead II in a manner inconsistent with the other two leads so that ventricular activation would produce positive deflections in all three leads of most normal records Differences in potential are recorded in the three limb leads according to the following relationship I = VL − VR, II = VF − VR, III = VF − VL, where VR, VL, VF represent potentials at the right arm, left arm, and left leg, respectively. Since the three leads form a closed circuit the algebraic sum of their potential differences at any instant is zero Because of the reversed polarity of lead II, this relationship (Einthoven's "law") is actually I − II + III = 0, or I + III = II This is really merely a consequence of Kirchhoff's second law of circuits The validity of this equation does not depend on any assumptions relating to the electrical homogeneity and geometry of the volume conductor or location of the heart

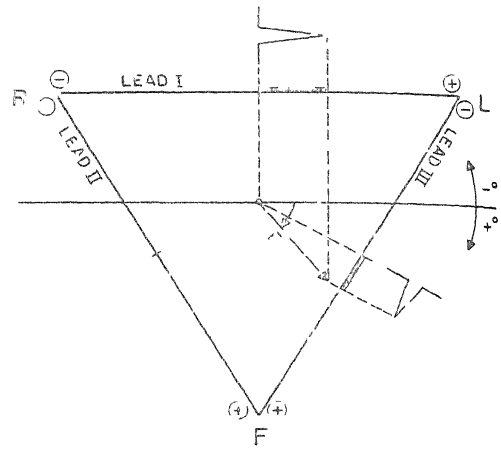


Fig 3 65 Einthoven triangle R right arm L left arm F foot Mean electrical axis is derived from lead I and lead III QRS complexes by drawing perpendiculars from the lead lines at points determined by amplitude of QRS Intersection of perpendiculars determines the mean electrical axis E

It is possible to derive the potentials at each extremity from these bipolar leads if Einthoven's hypotheses are correct That is, if the location of the extremity electrodes is equivalent to the apices of an equilateral triangle on the surface of a homogeneous volume conductor with a point source of electrical potential in the center, then VR + VL + VF = 0, and it can be demonstrated algebraically that VR = −(I + II)/3, VL = (I − III)/3, and VF = (II + III)/3 However, if Einthoven's hypotheses are correct it is also possible to obtain an indifferent electrode of zero potential from a network connecting the three extremities, and then one may determine the potentials at any point in the body with an exploring electrode Wilson (1934) demonstrated that the potential (VT) at a central terminal connecting the three extremities through equal resistances is equal to the mean of the potentials at the extremities, and so VT = (VR + VL + VF)/3 = 0 in this special situation Differences between resistances at the three skin-electrode interfaces which

would affect VT are minimized by a high input resistance in the circuits of modern electrocardiographic instruments Although Einthoven's hypotheses are only a first approximation of actual conditions, several studies suggest that changes in VT are so small during the cardiac cycle that they can usually be disregarded for most clinical purposes, and the central terminal is considered to have a constant or "zero" potential Tracings which record differences in potential between an exploring electrode and the central terminal are called "V" leads and are often described as "unipolar leads" to indicate that they essentially represent changes in potential at the site of the exploring electrode The exploring electrode in V leads is always attached to the positive pole of the electrocardiograph so that when a positive charge faces the exploring electrode a positive (upward) deflection is recorded Since the central terminal is considered to have a zero potential, the lead lines of V leads extend from the exploring electrode to the zero point at the center of the dipole, and projections of vector forces on these lines can be analyzed as described previously

Clinical electrocardiograms usually include six V chest leads recorded with the exploring electrode at the positions indicated in Figure 3 66 The normal precordial lead V_1 displays a small R wave due to the initial activation of the septum from left to right The small R is followed by a deep S wave due to activation of the free wall of the left ventricle which produces electrical forces directed to the left and away from this electrode Leads V_5 and V_6 are more or less the inverse of V_1 They usually show a small Q wave followed by a large R wave In the intermediate leads progressing from V_2 to V_5 there is increasing R wave amplitude and decreasing S amplitude Usually at V_3 there is approximately equal R and S amplitude This is referred to as the transition complex Occasionally the transition complex may be recorded as far to the right as the V_2 position or as far to the left as V_5

Goldberger (1947) suggested that the potential at an extremity could be obtained simply by disconnecting one extremity electrode from the central terminal and recording the difference in potential between it and the remaining two electrodes Such tracings have the same configuration and 1½ times the amplitude of a V lead recorded with an exploring electrode on the same extremity The Goldberger arrangement for recording "augmented unipolar" extremity leads (named aVR, aVL, and aVF) is incorporated into the lead selector switch of modern electrocardiograph instruments

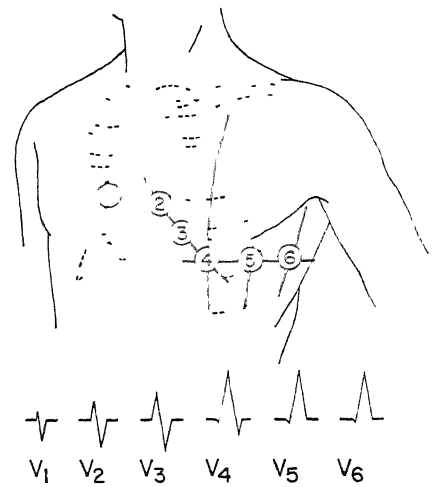


Fig 3 66 Positions of the chest electrodes for 6 standard precordial leads V_1 fourth intercostal space to right of sternum V_2 fourth intercostal space at left of sternum V_4 fifth interspace at midclavicular line V_3 midway between V_2 and V_4 V_5 anterior axillary line same level as V_4 V_6 midaxillary line same level as V_4 and V_5 Below are shown general form of QRS in normal precordial leads V_3 represents transition from predominant negativity on the right to predominant positivity on left (Lipman and Massie 1959)

Knowledge about the pathway of activation of the mammalian heart and the geometry of lead lines as developed from the Einthoven hypotheses can be used schematically to describe the effects of cardiac activation on the extremity leads Ventricular activation can be described by a series of instantaneous vectors, each of which is the resultant of dipoles representing the polarized surfaces present at that instant It is usual to consider only three or four such instantaneous resultant vectors (Fig 3 67) *1*) Initial activation of the septum is represented by a small vector directed to the right and forward *2*) Simultaneous activation of the lower septum, free wall of the right ventricle and anterior left ventricle is represented by a resultant vector directed to the left, forward, and downward *3*) Continuation of the activation process in the posterolateral left ventricular wall after its completion in most regions of the right ventricle is next represented by a large vector directed to the left and backward *4*) Terminal activation of the posterobasal left

ventricle and upper septum is represented by a small vector directed backward and upward

The interval between these instantaneous vectors is 0 01 to 0 015 sec in the dog heart and approximately twice as long in the human heart One can roughly predict the configuration of tracings in various leads by projecting these hypothetical vectors onto the appropriate lead line, or conversely one may derive the instantaneous vectors from the recorded tracings For example, potentials recorded at the same instant in two standard limb leads are measured and plotted in arbitrary units along the appropriate lead lines of an equilateral triangle, starting from the zero point at the center of each bipolar lead line and extending toward the end with the proper polarity Perpendiculars dropped from the ends of these two linear plots will meet at points which define the origin and terminus of the instantaneous mean vector responsible for that particular deflection (Fig 3 65) A single vector depicting the mean direction of the total electrical forces developed during ventricular activation can be derived by plotting the algebraic sum of positive and negative deflections in QRS complexes from each of two leads on the appropriate lead lines and dropping perpendiculars as described above The similarity between this derived vector, called the "mean electrical axis," and the true mean cardiac vector depends on the validity of the Einthoven hypotheses The electrical axis is described according to a system of polar coordinates with 0° at the right hand end of a horizontal line, positive values are below and negative values above this line (Fig 3 65) In normal humans, the electrical axis commonly falls between 0° and +90°, displacement in a clockwise direction beyond +90° is called right axis deviation, and displacement in a counterclockwise direction beyond 0° is called left axis deviation The electrical axis is influenced by the position of the heart, changes in the functioning myocardial mass or the pathway of activation, and changes in the conducting medium surrounding the heart A change in the direction of the long axis of the heart produces a shift in electrical axis in the same direction The electrical axis will shift toward an area of hypertrophied myocardium and away from an area of infarcted myocardium Since the electrical axis is determined from potentials recorded at the body surface, it will also be influenced by changes in the conducting medium, such as the introduction of air or fluid into the pleural space The electrical axis determined from the limb leads represents the projection of spatial vectorial forces onto a frontal plane defined by the Einthoven triangle, an additional lead having a lead line in an anteroposterior direction is required to estimate the spatial orientation of these forces

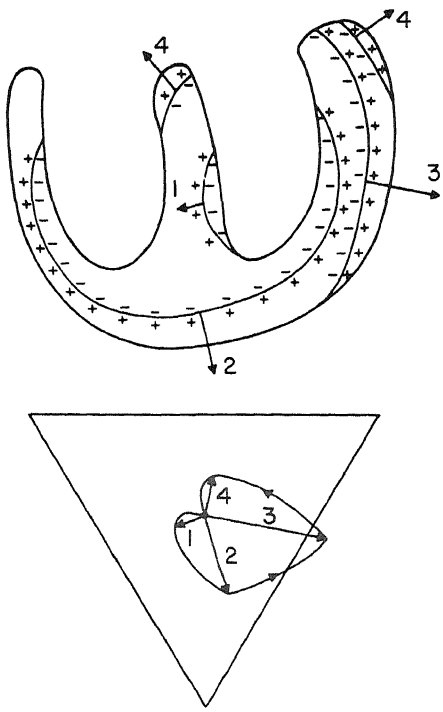


Fig 3 67 The four major dipole vectors generated during ventricular activation giving rise to the deflections of the QRS complex By Einthoven s hypothesis these vectors may all be considered as originating from a single point at the center of an equilateral triangle When the termini of all resultant instantaneous vectors are connected a loop is described The loop is the vectorcardiogram

A line joining the termini of all the instantaneous resultant vectors will form a loop representing the sequential change in amplitude and direction of the resultant electrical forces throughout the cardiac cycle (Fig 3 67) Such a loop can be derived by graphic or electronic integration of potentials from two simultaneously recorded leads When a loop is obtained by electronic integration on an oscilloscope, it is called a vectorcardiogram

RECOVERY PROCESS AND THE VENTRICULAR GRADIENT

The recovery process, which is simply a reversal of changes that occur during activation, is characterized by deflections of much longer duration and lower amplitude than the activation process Atrial recovery produces a prolonged deflection (T_p or T_a wave) which continues through the PR interval and into or beyond the QRS complex, but it is generally of such low amplitude that it is clearly defined only in tracings where P waves are widely separated from the QRS complexes, as in heart block The T_p is opposite in direction to the P wave in tracings obtained from strips of atrial muscle and also from the intact heart, which indicates that the recovery process follows the same sequence of activation since the two processes have opposite electrical signs Ventricular recovery occurs during inscription of the ST segment and T wave The ST segment corresponds to the plateau and the T wave corresponds to the rapid final phase of repolarization recorded from single myocardial fibers In man the tracing generally shows little or no deflection from the base line during inscription of the ST segment when large areas of myocardium are undergoing recovery simultaneously The T wave occurs when recovery is complete in some areas but is continuing in other regions The T wave recorded from muscle strips is opposite in direction to the QRS, as expected when recovery occurs in the same sequence as activation, but this is not true in direct leads on the surface of the intact ventricle where the T wave is in the same direction as the major deflection of the QRS This indicates that the sequence for completion of recovery in the intact ventricle is opposite in direction to the activation process, and therefore occurs later at endocardium than at epicardium

The relationship between the orientation of the electrical forces during activation and recovery has been expressed in vector terms by Wilson (1954) and others These investigators determined the net area enclosed by QRS deflections and that enclosed by ST and T deflections in the limb leads, and used these areas (which represent the product of time and voltage) to plot mean vectors called $\hat{A}$ QRS and $\hat{A}$ T, respectively The resultant of these two vectors is called the ventricular gradient $\hat{G}$ If $\hat{A}$ T was equal in amplitude and opposite in direction to $\hat{A}$ QRS, as expected on a theoretical basis from electrograms of muscle strips, their resultant or ventricular gradient would be zero The degree to which the actual $\hat{A}$ T deviates from the theoretical one is expressed by the vector $\hat{G}$ The gradient is thus a reflection of the difference between the sequence of activation and recovery

Abnormalities of the Electrocardiogram

VENTRICULAR HYPERTROPHY

Left Ventricular Hypertrophy This often causes the following electrocardiographic changes *1*) increased voltage in leads from the precordium, *2*) increased duration of the QRS complex, *3*) rotation of the mean electrical axis toward the left, and *4*) secondary ST and T wave abnormalities in left precordial leads (Fig 3 68) In the frontal plane the mean electrical axis comes to lie in the 0 to $-90°$ quadrant (left axis deviation) *Right ventricular hypertrophy* shifts the mean electrical QRS axis in the frontal plane toward the right (usually greater than $+105°$) The R wave in V_1 usually exceeds S in amplitude and the transitional changes in QRS configuration differ from normal in that there is a progressive decrease in R amplitude and an increase in S amplitude (compare Figs 3 66 and 3 68, top)

RIGHT BUNDLE BRANCH BLOCK

It will be recalled that in the normal sequence of activation the left side of the septum starts to depolarize slightly before the right side, the free wall of the right ventricle is depolarized simultaneously with the left ventricle, and the greater thickness of the left ventricle produces much larger voltages than those from the right ventricle Thus, the only rightward directed voltages seen in the normal electrocardiogram are those generated initially from the left side of the septum and in some instances a small late voltage from a late activation of the pulmonary conus

When the right conducting bundle is blocked either by cutting it experimentally or by some pathological process, the activation of the right ventricular free wall is greatly delayed (Fig 3 69) The first portion of the QRS com-

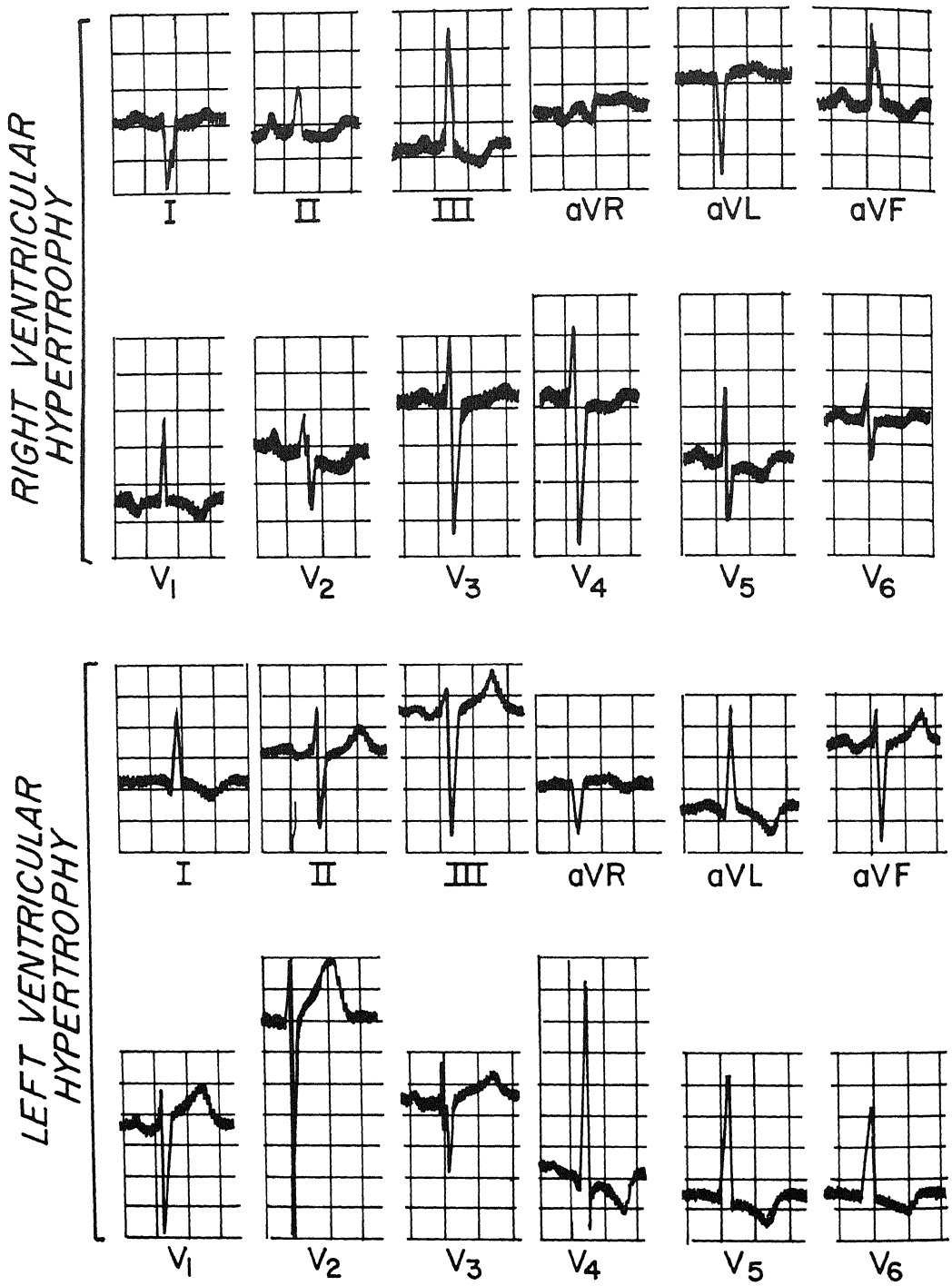


Fig 3 68 Electrocardiogram patterns of right and left ventricular hypertrophy

plex up to about 0 04 sec is inscribed in a normal manner as it records the activation of the septum from left to right and the free wall of the left ventricle When the left ventricle has become completely depolarized, there still remains a considerable portion of the right ventricular free wall which has not been depolarized The impulse may spread slowly in a

tangential direction through the free wall of the right ventricle rather than being transmitted almost instantaneously to all parts of the endocardium by Purkinje fibers and spreading from endocardium to epicardium This slow late right ventricular activation wave, now unopposed by left ventricular forces and not having the high proportion of internal cancella-

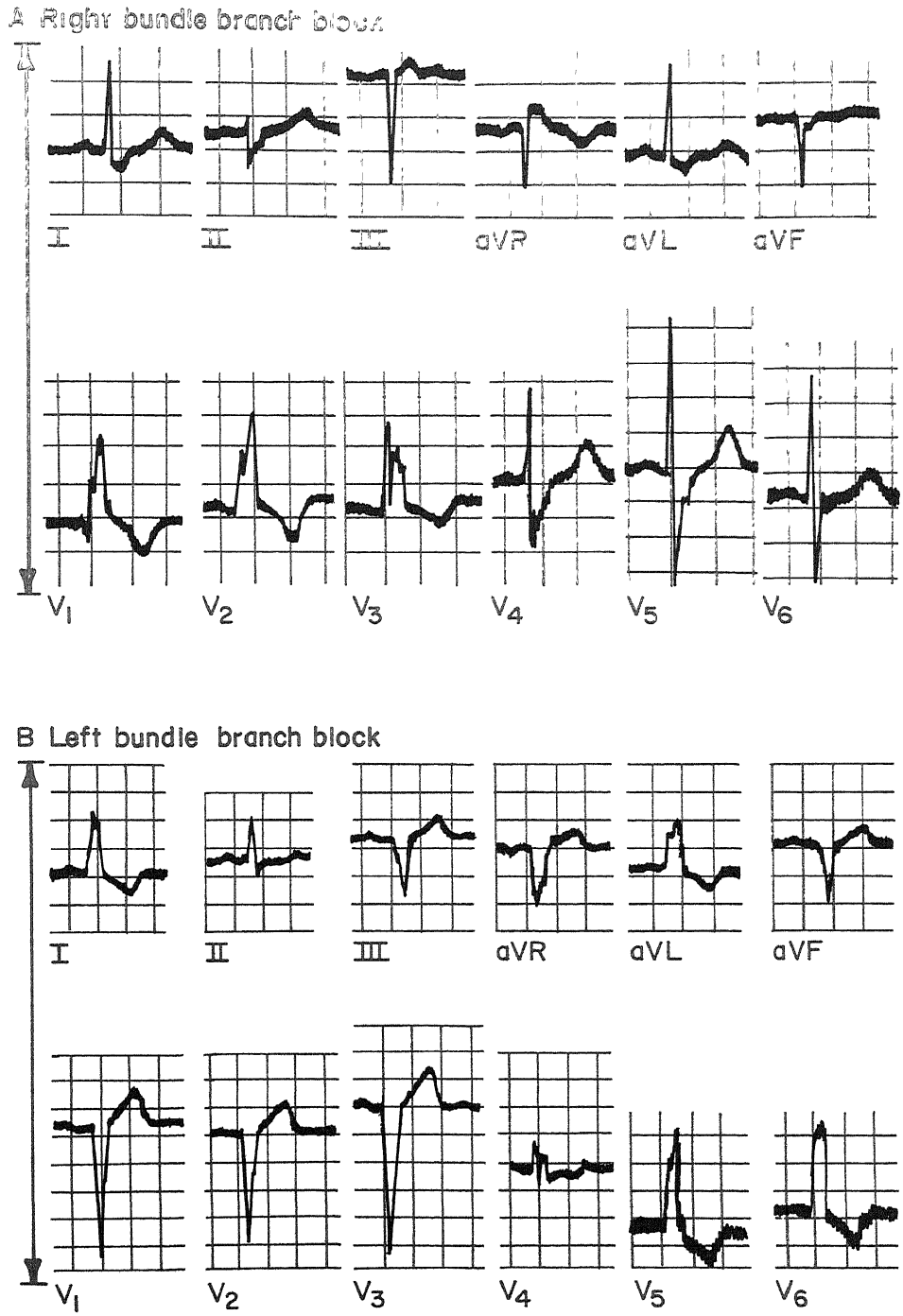


Fig 3 69 Patterns of right and left bundle branch block

tion characteristic of normal transmission, produces a large voltage directed anteriorly and to the right which appears late and prolongs the QRS to a duration of 0 12 sec or more

This late right ventricular voltage is inscribed in the electrocardiogram as a late slurred R wave in V_1 and aVR, and a late slurred S wave in V_5, V_6, and aVL

LEFT BUNDLE BRANCH BLOCK

Interruption of the left conducting bundle causes delay in activation of the left ventricle The first portion of the heart to be activated is now the right side of the septum by way of the right bundle branch The septum is thus activated from right to left, producing a large voltage directed leftward This large leftward voltage is opposed to a small degree by a rightward voltage from the almost simultaneous activation of the free wall of the right ventricle, but because of the greater muscle mass of the septum the net initial voltage is directed toward the left The effect on the electrocardiogram is to produce initial positivity in leads I, V_5, and V_6 The normal septal Q wave is thus abolished in these leads

The wave of activation is transmitted through the septum to the free wall of the left ventricle through which it passes slowly in a tangential direction from apex to base It is probable that there is also some transmission from the anterior and posterior septum circumferentially so as to envelope the free wall in the activation wave The slow activation of the left ventricle continues to produce leftward directed voltages which are inscribed in the electrocardiogram as broad notched or slurred R waves in the left precordial leads V_5 and V_6 Duration of the QRS complex is 0 12 sec or greater In general the ST and T waves are opposite in direction to the QRS complexes

MYOCARDIAL ISCHEMIA, INJURY, AND NECROSIS

When a coronary artery is gradually compressed, the first change seen in tracings from the area of myocardium supplied by that vessel is a deep and symmetrical inversion of the T wave This symmetrical form of T wave inversion is referred to as T wave inversion of the ischemic type (Fig 3 70) It is believed to be due to delay in recovery at the epicardial region so that there is reversal of the normal order of repolarization

The ischemic T wave inversion is rapidly reversible if adequate circulation is restored to the myocardium When the blood supply to the myocardium is more drastically impaired by sudden ligation of a coronary artery, the direct epicardial tracing shows the rapid onset of elevation of the ST segment This elevation may be almost as high as the QRS complex and may yield a wave form closely resembling the transmembrane action potential obtained with micropuncture of a single myocardial fiber The ST elevation is believed to be due to a strong current of injury causing marked depression of the resting isopotential line The ST displacement may be accompanied by ischemic changes in the T waves This injury pattern is also rapidly reversible upon restoration of adequate blood flow

When a coronary artery is permanently ligated, a portion of the myocardium becomes infarcted The presence of electrically inert necrotic myocardium results in alterations of the QRS complexes of the direct epicardial lead If the infarction is completely transmural from epicardium to endocardium this mass of tissue contributes no potential of its own, but simply conducts to the surface the potentials which are to be found in the underlying cavity The depolarization wave in this case is a purely negative QS complex identical to that found in the left ventricular cavity The repolarization phases contain elements of injury and ischemia patterns derived from adjacent viable portions of the myocardium

Exploration of the adjacent areas of myocardium about a zone of necrosis discloses a zone of injury surrounding the necrotic areas (Fig 3 70) Complexes recorded from the injury zone are characterized by marked elevation of the ST segment without significant change in the morphology of the early portion of the QRS The injury zone is surrounded by a zone in which the T waves show inversion of the ischemic type This is called the ischemic zone Beyond it lies normal muscle disclosing normal QRS-T complexes

Electrocardiograms made from the surface of the body include potentials from the entire heart, and therefore, the precise localization of zones of necrosis, injury, and ischemia cannot be derived from the clinical electrocardiogram When infarction is present, all of the phenomena observed in the three zones will be reflected in the appropriate electrocardiographic lead Even a precordial lead adjacent to a large

area of infarction is influenced by electrical events in other parts of the heart and is likely to record elements from all three of the zones of necrosis, injury, and ischemia

Thus, the classical pattern of myocardial infarction consists of a deep wide Q wave, ST segment elevation, and T wave inversion Muscle necrosis cannot be identified with certainty from the electrocardiogram without the appearance of a pathological Q wave which indicates a removal of electrically active myocardium ST and T wave changes of injury and ischemia may occur when myocardium becomes temporarily or even chronically hypoxic but not sufficiently so to produce muscle death On the other hand, there may be myocardial necrosis in areas completely surrounded by viable tissue or in subendocardial zones such that the general pathway of activation might not be markedly altered In such intramural or subendocardial infarctions abnormal Q waves may not appear even though injury and ischemic ST and T changes may be present In such cases the presence of muscle necrosis cannot be proven by the electrocardiogram Evidence of its presence might, however, be obtained by other clinical means

Localization of Myocardial Infarction. The general anatomical localization of areas of cardiac infarction may often be made from the standard 9 or 12 lead electrocardiogram Each lead constitutes, so to speak, an electrical view of the heart from a different point in space (Fig 3 71)

DRUGS AND ELECTROLYTES

Digitalis produces a characteristic alteration of the ST and T complex which might sometimes be mistaken for an injury effect (Fig 3 72) The change consists of a shortening of the QT interval and a sagging of the ST segment The sagging ST has a "scooped-out" or "cuplike" concavity that is easily recognized This electrocardiographic sign of digitalis effect is not related to the adequacy of the dose of the medication nor to drug intoxication

Quinidine and procaine amide have similar effects on the electrocardiogram They cause depression of the ST segment, a decrease in amplitude or diphasic configuration of the T wave, and an increase in amplitude of the U wave The QT interval may be prolonged, or may appear prolonged because the U wave is superimposed on the descending limb of the T wave producing a broad T-U complex which may have a notched or undulating appearance As the dose is increased slowing of conduction

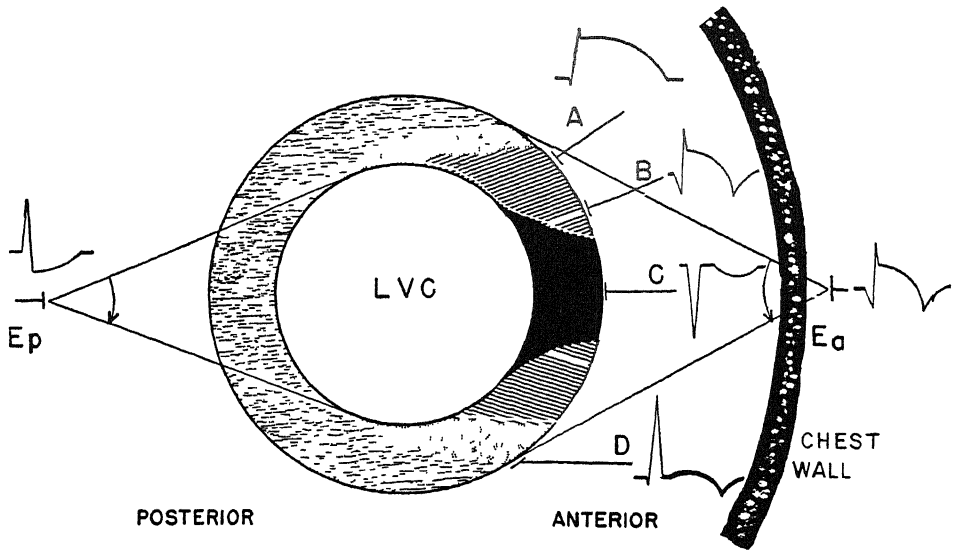


Fig 3 70 Schematic diagram of direct epicardial leads taken over an area of myocardial infarction and over adjacent zones of injury and ischemia Also shown is a chest lead facing the infarct and a lead from the opposite side of the heart At *C* cavity negativity is reflected in the epicardial lead over the necrotic zone At *A* is recorded a monophasic injury wave over the injured zone At *B* the complex reflects a combination of necrosis and injury At *D* there is only T wave inversion of the ischemic type over the ischemic zone At E_a over the chest the complex combines features of all three zones and on the opposite side of the heart there are reciprocal changes of increased positivity of the R wave and depression of ST (From Barker 1952)

through the ventricle is produced with resulting widening of the QRS complexes

Among electrolyte ions, calcium and potassium have the most significant effects on the electrocardiogram (ECG) Hypocalcemia causes prolongation of the QT interval due to prolongation of the ST segment without distortion of the T wave The long flat ST segment is followed by a T wave of normal amplitude and duration Micropuncture studies of transmembrane action potentials have shown that the long ST segment is due to prolongation of the plateau phase of the recovery process The later phases continue at their usual rates

Elevation of the serum calcium has the opposite effect of shortening the ST segment by shortening the plateau phase of the monophasic action potential The T wave itself appears undisturbed

Reduction of serum potassium appears to have two effects, one on the ST-T complex and one on the U wave When the potassium level

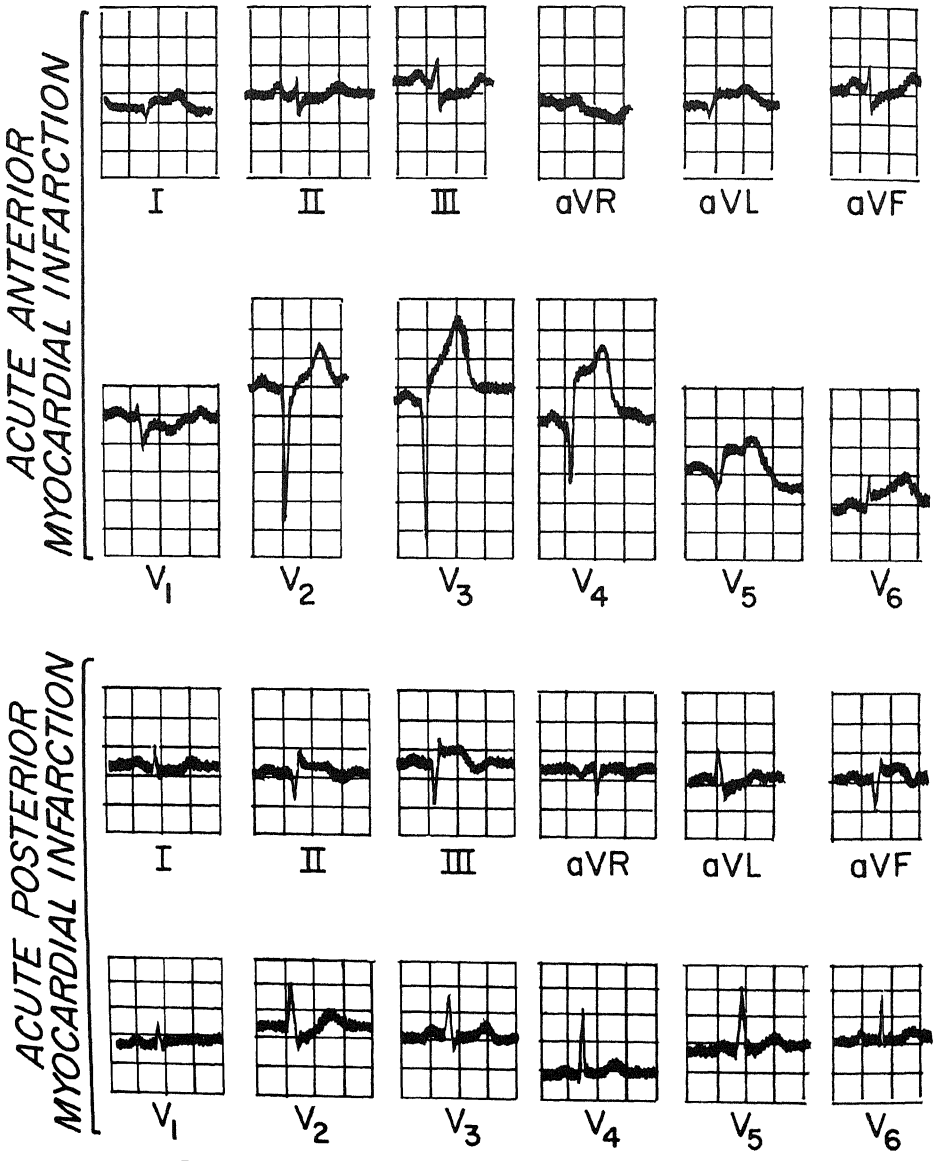


Fig 3 71 Patterns of acute anterior and posterior myocardial infarcts

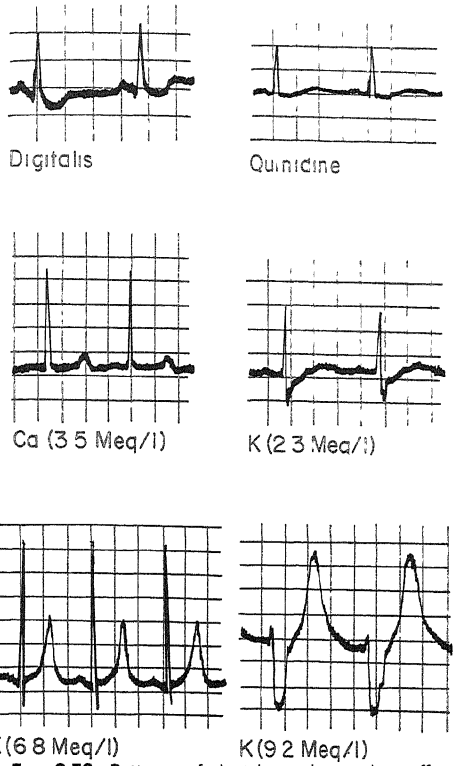


Fig 3 72 Patterns of digitalis and quinidine effect low serum calcium low serum potassium and high serum potassium

in the serum falls to about 3 mEq /liter, the ST begins to sag and the T wave becomes lower while the U wave becomes more prominent The T and U seem to merge into a single wave separated only by a slight indentation or dimple on the crest This gives a false impression of a markedly prolonged QT With very low levels of potassium the ST segment becomes depressed below the base line and the U wave exceeds the T in height This gives the appearance of a stepwise ascent of the ST-T complex This pattern is usually seen with serum potassium of 2 5 mEq /liter or less

Elevation of serum potassium leads first to an increasing amplitude of the T wave in leads where the T is normally upright The tall T waves have a tent-like appearance with similar slopes on the upward and downward limbs With higher levels of serum K (8 mEq /liter approximately), the PR interval becomes prolonged and the QRS begins to widen out as conduction becomes impaired At about 9 mEq /liter, the P wave disappears entirely and the QRS takes on a bizarre shape merging with the tall T wave to produce a distorted sine wave-like pattern which usually heralds terminal ventricular fibrillation or standstill

Cardiac Arrhythmias and Conduction Disturbances

The following is a convenient classification of cardiac irregularities

- A Affections of rhythm due to impaired conduction through the AV node and the bundle of His
  - I Delayed conduction (concealed conduction)
  - II Missed beats, partial heart block
  - III Complete heart block
- B Wolff-Parkinson-White syndrome
- C Affections due to abnormal impulse formation
  - I Extrasystoles
    - (1) Ventricular
    - (2) AV junction
    - (3) Atrial
  - II Paroxysmal tachycardia
    - Atrial, AV junctional, and ventricular
  - III Atrial flutter
  - IV Atrial fibrillation
  - V Ventricular fibrillation
- D Alternation of the heart
- E Affections due to vagal influences
  - I Sinus arrhythmia
  - II Phasic irregularity
  - III Sinus bradycardia
  - IV Sinoatrial block
  - V Atrioventricular block

AFFECTIONS OF RHYTHM RESULTING FROM IMPAIRED CONDUCTION

In the AV Node or Stem of the Bundle Atrioventricular Block. In animals, conduction from atrium to ventricle can be depressed or blocked by crushing, cutting, or the application of cold to the AV bundle This strategic point in the pathway of the excitation wave is also attacked by disease, and conduction through it may be depressed or completely abolished

Concealed conduction is a term used for the situation in which impulses penetrate the AV conduction system but fail to emerge from it (Moore et al , 1971) In the electrocardiogram

such concealed impulses are inferred owing to their known effects on subsequent characteristics of AV conduction such as *1*) delay or blocking of AV conduction of a succeeding impulse, *2*) acceleration of AV conduction of a succeeding impulse, *3*) delay or abolition of the expected discharge of an ectopic pacemaker located in AV junctional tissues Concealment and its subsequent effects are physiological phenomena related to the conduction and recovery characteristics of the AV conduction tissues

Depression of conduction through the node or bundle varies in degree Three stages are recognized

Delayed Conduction (Incomplete Atrioventricular Block) In this stage conduction is merely slowed, every impulse reaches the ventricle The intervals between the atrial and ventricular systoles (A_s-V_s intervals) are lengthened beyond the normal maximum of 0 2 sec and may have a duration of 0 5 sec , though, as a rule, they are considerably shorter than this The condition can be recognized by means of the electrocardiograph or from a venous pulse tracing Lengthening of the PR interval in the former tracing, or of the a-c interval in the latter, beyond 0 2 sec is taken to indicate delayed conduction (Fig 3 73)

Missed Beats—Partial Heart Block (Incomplete Atrioventricular Block with Dropped Beats) When impaired conduction reaches a certain degree, impulses from time to time fail to reach the ventricle, and a beat is missed The atrial beats are perfectly regular, and in this way the condition differs from sinoatrial block A ventricular beat may be missed only occasionally and at irregular intervals The PR interval of the electrocardiogram, or the a-c interval of the venous pulse, may generally, though not invariably, lengthen progressively for several heart cycles preceding a dropped beat The periods of progressively delayed conduction preceding the missed beat are called, after their discoverer, the Wenckebach periods The interval of the cycle succeeding the missed beat is shortened again to near the normal length In a further stage of the condition the beats are dropped more frequently and may be spaced at either regular or irregular intervals in the tracing When the grade of block is still more advanced, impulses fail to penetrate the bundle after every second atrial beat, or three, or even four atrial contractions may occur before an impulse reaches the ventricle, i e , the ventricle responds only to every third or fourth atrial beat So, an atrioventricular rhythm becomes established in which the two chambers beat in the ratio of 2 1, 3 1, or 4 1

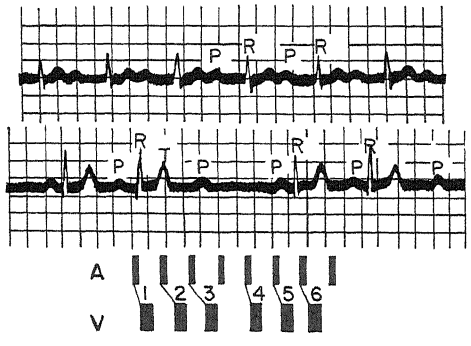


Fig 3 73 Upper tracing delayed conduction lead I Lower tracing incomplete heart block lead I (missed beats) (Kindness of Dr John Hepburn) The diagram below (after Lewis) represents incomplete heartblock The thin rectangles *A*, represent contractions of the atria the thicker ones, *V* contractions of the ventricle The obliquely directed lines represent conduction over the AV bundle the slower the conduction rate the more oblique the line The gaps in the lower rectangles indicate missed beats of the ventricle It will be noted that delay in conduction increases progressively in successive cycles until a beat is missed Heavy vertical lines = ⅓ sec

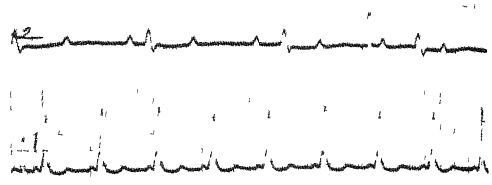


Fig 3 74 Above—complete heart block atrial rate 75 ventricular rate 36 lead 2 Below—Wolff Parkinson-White syndrome lead 1 PR interval 0 09 sec QRS interval 0 11 sec

Complete Heart Block (Complete Atrioventricular Block) When the AV node or bundle offers an absolute barrier to the passage of the impulse, the dissociation of the rhythms of the upper and lower chambers is complete (Fig 3 74) The atria beat at their own rate of about 70/min and the ventricles at their inherent rate of about 35 The impulses that drive the ventricles are believed to arise only in the AV junctional region, bundle of His, right or left bundle branches or in the Purkinje fibers Hoffman has shown through extensive investigation of the transmembrane action potentials of many areas of the heart that only the special-

ized conducting tissues have the property of automaticity that would permit their cells to take over pacemaking function The property of automaticity is characterized by the ability of the cell to undergo gradual spontaneous diastolic depolarization to the threshold level of potential at which an action potential spike is initiated Ventricular myocardial cells do not have this property The speed with which a particular region of the heart can develop and discharge impulses apparently determines its ability to dominate other regions It has been mentioned that when a region of higher rhythmicity is destroyed or isolated the region next in order of rhythmical power assumes the role of pacemaker When, for instance, the SA node is destroyed or isolated the AV junctional region assumes control, and when this or the upper part of the bundle is separated from the tract of tissue below, the lower lying portion takes over the government of the ventricular rate

WOLFF-PARKINSON-WHITE (WPW) SYNDROME (ACCELERATED CONDUCTION, VENTRICULAR PRE-EXCITATION)

In this rare congenital anomaly the PR interval of the electrocardiogram is shortened and the QRS complex prolonged The PR interval is 0 12 sec or less and the QRS complex extended to over 0 10 sec (Fig 3 74) The shortened PR interval may be due, in some cases, to an accessory conducting strand between atrium and ventricle Butterworth and Poindexter (1942) fashioned an extra electrical conduction pathway in the dog's heart, which, when stimulated, produced an electrocardiogram similar to that characteristic of the syndrome Wood et al (1943) in the histological examination of the heart of a patient which had shown this syndrome, also found three accessory pathways from the right atrium to the ventricle

Moreover, Boineau, and Moore (1970) demonstrated a physiologically active atrioventricular muscular bridge (Kent bundle) accounting for ventricular pre-excitation in a dog with spontaneous WPW Further, surgical separation of the right atrium and right ventrical or digital pressure at the site of ventricular pre-excitation in humans with WPW have been shown to normalize atrioventricular conduction Prinzmetal and his colleagues (1952) have produced typical WPW complexes in dogs by applying a noninterrupted direct current of subthreshold strength to the AV node, but such complexes could not be induced after severance of the AV bundle—an indication that the impulses traverse the normal AV connections These investigators carried out a large number of similar experiments in which the movements of the heart were recorded by high speed cinematography, and the records analyzed after slowing down to half speed They conclude from their observations that the accelerated conduction of the WPW syndrome is due to shortening of the delay of the impulse at the AV node to a fraction of the normal which amounts to about 0 12 sec

These experiments indicate the WPW syndrome can have a physiological as well as an anatomical basis

The prolongation of the QRS complex is attributed to early activation of a portion of myocardium followed by normal activation of the remaining muscle The significance of this unusual mechanism lies in the fact that persons who display this electrocardiographic pattern are particularly prone to have frequent bouts of paroxysmal rapid heart action of one type or another

DISTURBANCES OF RHYTHM DUE TO ABNORMAL IMPULSE FORMATION

Extrasystoles or Premature Contractions. An extrasystole can be induced experimentally by stimulating the cardiac muscle at any time except during its phase of absolute refractoriness Extrasystoles occur in the human heart as a result of some abnormal process of impulse formation Though extrasystoles may be associated with organic heart disease they more frequently occur in its absence, they may then be reflexly initiated from the abdominal viscera or be due to some form of intoxication, e g , digitalis, cyclopropane anesthesia, hyperthyroidism, excessive smoking, excessive coffee or alcohol ingestion, stimulant drugs such as amphetamine, nervous tension, anxiety, or exhaustion Electrical stimulation of various parts of the brain (e g , sigmoid gyrus, limbic lobe, hypothalamus, amygdala, etc) in cats can produce atrial and ventricular extrasystoles (Hoff et al , 1963), and their occurrence in man following brain lesions has been reported, which indicates that in some instances they are of central origin The atrium or the ventricle may

be the site of origin of the premature contraction, or the extra impulse may arise in the AV junctional tissues (see diagrams, Fig 3 75)

Ventricular Extrasystoles The premature contraction occurs after the normal ventricular beat has ceased and the muscle has recovered from its absolute refractory state It is not preceded by an atrial contraction, and is not dependent upon an impulse received from the upper chamber (Fig 3 76) Since the impulses arise in one or the other bundle branches or in the Purkinje fibers, the electrocardiographic patterns of ventricular premature beats often resemble those of right or left bundle branch block They are broad and bizarre not resembling the normal complexes The premature contraction is followed by a long pause This is

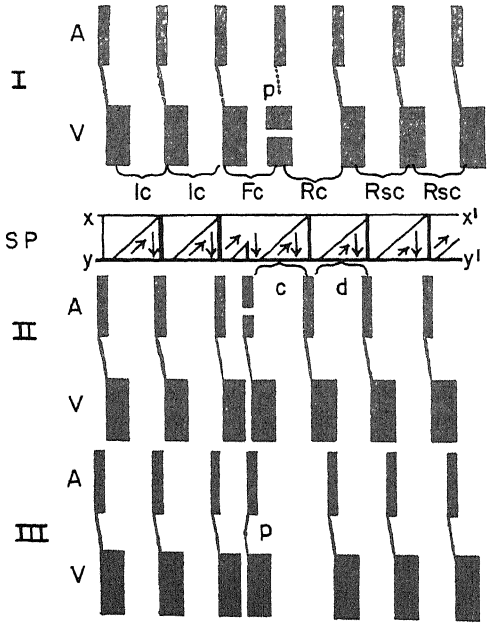


Fig 3 75 (After Lewis) *I* a diagram illustrating disturbance of the heart s mechanism when a systole is caused by exciting the ventricle during diastole Ic initial cycle Fc forced or extrasystolic cycle Rc returning cycle, and Rsc restored cycles p is the premature or forced beat Note that the atrial rhythm remains undisturbed The forced and returning cycles are together equal in length to two initial cycles *II* a diagram illustrating the events when a premature contraction is excited from the pacemaker Stimulus production in the tissue which originates the heart rhythm is indicated by the line S P, the impulse is supposed to discharge when it reaches the line xx' and to fall at each contraction of the heart to the level yy' c and d are equal in length *III*, a diagram illustrating a premature beat arising in the AV junctional tissue

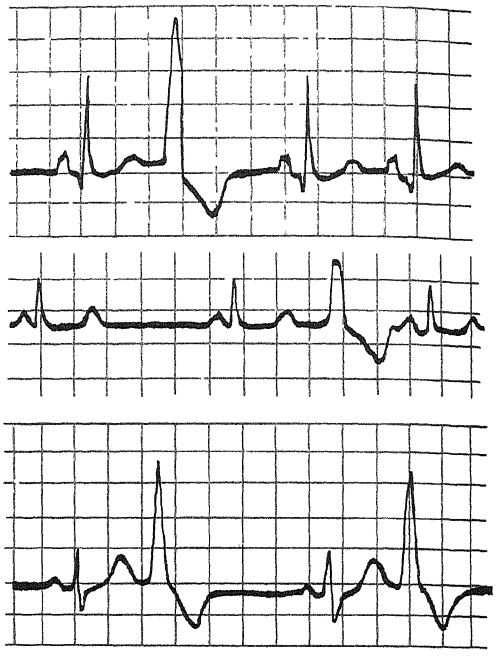


Fig 3 76 *Upper* ventricular extrasystole with compensatory pause *middle* interpolated ventricular extrasystole *lower* coupled ventricular extrasystoles (bigeminal rhythm) Note that the intervals between the ectopic ventricular beat (extrasystole) and the preceding normotopic beat are identical Such 'fixed coupling intervals are characteristic of ventricular ectopic beats arising from the same focus

usually of just sufficient duration to cause the succeeding normal ventricular beat to occur at the instant that it would have occurred had there been no premature contraction The cause of this *compensatory pause* has been explained elsewhere Briefly, it is due to the normal impulse reaching the ventricle when the muscle is still refractory as a result of the premature beat The interval between the normal beat and the one following the premature contraction is therefore equal in length to two normal cardiac cycles (Fig 3 75*I*) Sometimes, however, when the extrasystole occurs early in diastole and the heart rate is slow, there may be no compensatory pause The atrial impulse then reaches the ventricular muscle after it has recovered from the refractory state resulting from the premature beat, the atrial impulse therefore brings about a response at, or (as a result of some lengthening of the As-Vs interval following the premature beat) slightly after the usual time The normal ventricular systoles are then all equally or nearly equally spaced and

the extra contractions are interposed here and there between them That is, the time interval from the normal beat preceding the extra contraction to that following it is of normal length or but slightly lengthened Premature beats of this nature are called *interpolated extrasystoles* (Fig 3 76) It should be noted that except in the case of interpolated beats there is not an extra or additional beat, as the term extrasystole seems to imply The premature beat, in effect, displaces the normal beat

Atrial Extrasystoles The premature contraction arises in the atria at some point outside the SA node The abnormal impulse reaches the ventricle along the usual paths evoking a ventricular contraction and giving rise to an electrocardiographic complex similar to the normal ones If the atrial beat is so premature that sufficient time has not elapsed to permit complete recovery of the ventricular muscle from its refractory state, then the impulse may be conducted aberrantly through the ventricle giving rise to a weak contraction and to an electrocardiographic complex having a different pattern from the normal beats The premature atrial beat prevents the occurrence of the next normal atrial impulse and the pause of the atrium which follows the abnormal atrial contraction is usually precisely equal to a normal interval

AV Junctional Extrasystoles Extrasystoles occur sometimes as a result of impulse formation in automatic cells located in a specific region of the AV junction or supraventricular part (stem) of the bundle The terms AV rhythms or AV junctional rhythms have been adopted because there is uncertainty about their anatomical origin (Marriott, 1966) Spontaneously automatic or pacemaking cells are found in tissues around the AV node and in the His bundle, but not in the node proper (Hoffmann and Cranefield, 1964) On account of its central position between the two chambers, impulses arising in these cells pass upward and downward to cause simultaneous or nearly simultaneous responses from atria and ventricles Sometimes the ventricular contraction may actually occur first, in which case it is suggested that the impulse has arisen in the stem of the bundle and so has its course to the ventricular muscle considerably shortened The extra cycle is usually, though not invariably, followed by a compensatory pause (Fig 3 75*III*)

In rare instances extrasystoles arise as a result of abnormal impulses initiated in the sinoatrial node—*sinus extrasystoles* Except for the interposition of the extra beat the rhythm of atria and ventricles is but little disturbed The interval following the extra beat is normal in length or slightly shortened

Effect on the Arterial Pulse Several long-recognized irregularities of the pulse are the result of extrasystoles For example, in the irregularity known as *intermittent pulse* there appear from time to time relatively long intervals during which no beat is felt in the radial pulse (Fig 3 77*a*) The intervals are most pronounced when a premature contraction of the ventricle is too weak to open the semilunar valves When the extrasystoles are forceful enough to open these valves, and when they occur regularly one after each normal systole, paired pulse beats, each couple being followed by a long pause, are felt in the radial pulse (Fig 3 77*b*) This type of irregularity, sometimes seen following overdosage with digitalis, is called the *bigeminal pulse*

Paroxysmal Tachycardia. This may be defined as a condition in which the rate of the heart is greatly accelerated for a longer or shorter period without obvious cause The rate varies in different cases from 140 to 250/min The onset of the paroxysm is sudden and the increased rate is maintained for a variable length of time with perfect regularity, successive cycles usually not varying in length by more than a hundredth of a second The paroxysm lasts for only a few beats in some instances, in others it persists for a few minutes, hours or even days, though attacks of more than ten days' duration are very rare (Lewis, 1925) The attack ceases as abruptly as it commenced, the heart resuming its normal rate almost instantly

The paroxysm, it is believed, consists of a series of rapidly recurring extrasystoles which completely submerge the physiological rhythm

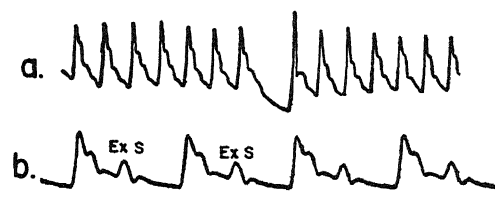


Fig 3 77 *a* Intermission of the pulse *b* pulsus bigeminus due to a single extrasystole with its succeeding compensatory pause occurring regularly after each normal beat *Ex S* extrasystole (After Price)

The site of origin of the extrasystoles, as in the case of single premature beats, may be in the atria, the *AV junctional tissues*, or in the *ventricle* (Fig 3 78) The atrial type is the most common, each atrial impulse spreads to the ventricle and causes a contraction whose features as indicated by the electrocardiogram are normal The P wave is frequently inverted

When the impulses arise in the AV junctional tissues, the PR intervals of the electrocardiogram are shortened Inversion of the P waves is common, or, the contractions of the two chambers may be simultaneous, the P waves being then buried in the ventricular (QRS) complexes Again, the contraction of the ventricle may occur before that of the atrium, it then sometimes happens that a progressive lengthening of the intervals between the R and P waves (RP interval) of the electrocardiogram is seen, ultimately an atrial beat is missed This is termed *reversed heart block* In other instances, as the RP intervals reach a certain length, a contraction of the ventricles occurs prematurely, and is not followed by a contraction of the atria It is thought that the ventricular contraction is caused by the same impulse that caused the preceding atrial beat That is, the impulse arising in the nodal region first excites the atrium, then re-enters the junctional tissue, which has now recovered from its refractory phase, and passes downwards to the ventricle This is spoken of as *reciprocal rhythm*

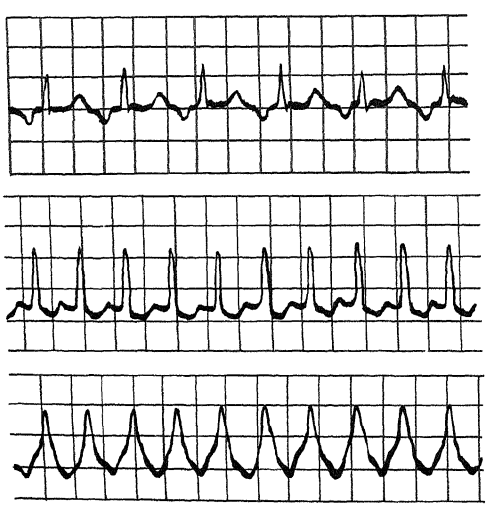


Fig 3 78 *Upper* nodal tachycardia at a rate of 120, lead 2 (note inverted P waves) *middle* atrial tachycardia at a rate of 200 lead 2 *lower* ventricular tachycardia at a rate of 200 lead 2

When the impulses arise in the ventricle the QRS complexes have the characteristics of those caused by ventricular extrasystoles The atrial rhythm is usually undisturbed, but occasionally it is abnormal, for when a series of rapidly recurring contractions arises in the ventricle, the impulses may pass along the bundle in a retrograde fashion and activate the atria to the exclusion of the normal impulse In other words, the ventricle then sets the pace, and the atria follow In such instances the P waves are inverted and succeed the QRS deflections, or are buried in the ventricular complexes as in the AV junctional type mentioned above

Atrial Flutter. There are two forms of this disorder, *pure* and *impure flutter* In both types the atria beat at the phenomenally rapid rate of from 250 to 400 beats per min , but in pure flutter the rhythm is regular, in impure flutter it is irregular Flutter differs from paroxysmal tachycardia in the following particulars *a*) the atrial rate of beating is usually much greater than that seen in paroxysmal tachycardia *b*) the disorder is of much longer duration, persisting unchanged for months or years, though it is sometimes transient *c*) the ventricles fail as a rule to follow the rate of the atria, a state of relative heart block becomes established as a result of the comparatively long refractory phase of the functional tissue

Atrial flutter may become converted to atrial fibrillation

In pure flutter the rhythm is remarkably regular (Fig 3 79) The lengths of the atrial cycles vary no more than a few thousandths of a second over comparatively long periods In impure flutter this constancy of cycle length is not seen and varying degrees of irregularity occur In flutter the atrial walls do not completely relax That is, though the proportion of active to inactive fibers varies during the rapid beating, at no one time are all the fibers in the relaxed state In the electrocardiogram, therefore, the level of the tracing between the P waves is continually changing and the electrical changes of the atria are carried over to alter the form of the ventricular complex

The ventricles rarely keep pace with the racing atria, the refractory phase of the con-

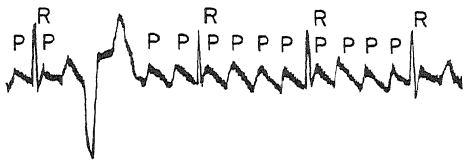


Fig 3 79 Pure flutter with a ventricular extrasystole (Kindness of Dr John Hepburn)

ducting tissue being longer as compared with that of the atrial muscle As a consequence, a state of relative heart block develops and an atrioventricular rhythm of 2 to 1, or less frequently, of 3 to 1 or 4 to 1, becomes established Sometimes there is variable block with 2 to 1, 3 to 1, 4 to 1, and occasionally even greater intervals occurring in the same strip of record Should the ventricle respond to each beat of the auricle, as occasionally happens, serious circulatory disturbances follow, ventricular diastole is so shortened that the ventricle receives a greatly reduced load of blood The output of the heart may be so reduced as a result that loss of consciousness resulting in death may follow

Atrial Fibrillation. In atrial fibrillation sequential coordinated excitation of the atrial musculature is replaced by chaotic, asynchronous contraction of small localized areas The atrial cavity is never emptied of blood and its wall is a quivering sheet of muscle Atrial fibrillation is the most common of all the serious cardiac irregularities It is most frequently seen in mitral stenosis and in thyrotoxicosis, but sometimes occurs in the absence of myocardial disease

Only a proportion of the atrial impulses pass through the AV bundle and activate the ventricle The relatively long refractory period of the conducting tissue shields the ventricle from the high rate of the atrial impulses The arterial pulse, nevertheless, is usually considerably faster than the normal (100 to 150) though it may be normal or even slowed Those impulses which reach the ventricle do so in a somewhat haphazard manner, and indeed one of the most characteristic features of fibrillation of the atria is absolute irregularity in the rate and force of the ventricular beats These features are expressed in the terms "*delirium cordis*," "*complete irregularity of the pulse*," or "*perpetual arrhythmia*," which were applied to the condition before its true nature was recognized A proportion of the heart beats are frequently so weak that they fail to cause pulse in the radial The apex beat is therefore much more rapid than the pulse The former, for example, may be 150 and the latter only 60 or 70 The difference is called the *pulse deficit* With treatment and improvement in the condition of the cardiac muscle, the pulse rate therefore may increase

The venous pulse in atrial fibrillation is of the ventricular form, *a* waves are absent, being represented by a series of rapid vibratory waves (*f* waves) Similarly in the electrocardiogram, small rapid undulations replace the *P* waves (Fig 3 80)

Processes Underlying the Production of Flutter and Fibrillation of the Atria *The Circus Movement Theory vs the Theory of Numerous Separate Points of Excitation* The observation that fibrillation could be induced in animals by electrical stimulation has led to a much clearer understanding of atrial fibrillation and flutter in man Lewis (1925), employing faradization as the method of inducing fibrillation, studied the subject intensively and, applying the results of the fundamental work of Mayer (1908), of Mines (1913), and of Garrey (1924), conceived that the excitation wave followed a circular pathway through the atrial muscle—the so-called *circus movement* That is to say, the wave starting at one place took a devious course through the cardiac musculature, returned to the point from which it started and re-entered the path which it had previously traversed Mayer induced a circus movement in the umbrella of the jellyfish (Medusa) by creating a local block and applying a stimulus to one side of the blocked region (Fig 3 81 top) The contraction wave which resulted was forced as a result of the block to take a unidirectional course, and after completing the circuit of the disk of tissue returned to the region of the block, which by this time had disappeared If the tissue from which the wave had been initiated was again excitable, i e , had passed from the refractory state, the wave circled the ring a second time, then a third time, and so on repeatedly When, on the other hand, the disk was stimulated in the absence of a block, a contraction wave set out in both directions and the two waves meeting, after having completed half the circumference of the strip, were suppressed at *B* (Fig 3 81 *bottom*)

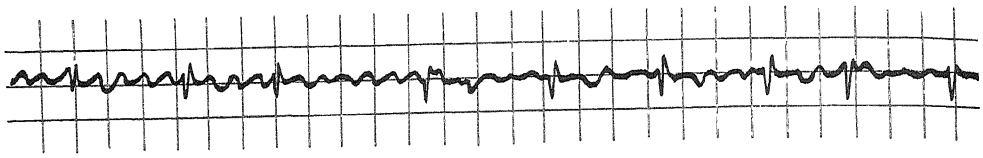


Fig 3 80 Atrial fibrillation lead V_1

That is, further progress of the waves was arrested, for each came to a region of tissue which, being occupied by the other wave, was refractory

Mines and Garrey linked up Mayer's observations upon Medusa with the fibrillation of cardiac muscle Mines showed that a similar circular motion of the contraction wave could be induced in the cold-blooded heart Muscular rings cut from the atria of large rays were employed Garrey thought that he had refuted the prevailing belief at that time, namely, that fibrillation of the atrium was due to the initiation of impulses from multiple foci throughout the auricular muscle—as proposed originally by McWilliam (1887) Garrey (1924) found that if he cut a small piece from the fibrillating atrium, the severed tissue at once ceased to fibrillate, an event which he contended could not occur if numerous points of excitation existed, but such would naturally result if the impulses travelled over a circular pathway from which other parts of the auricle were excited He induced a circulating contraction wave in large annular strips cut from turtles' ventricles

Evidence was obtained by Lewis (1925) which seemed to show that the impulses in the fibrillating atrium travelled in the natural rings of muscle encircling the openings of the venae cavae

The theory of circus movement has been very widely though not universally accepted Brams and Katz (1931) contended that if a circus movement is responsible for flutter and fibrillation more than one circulating wave must exist They base this conclusion upon the observation that if after experimental flutter or fibrillation has been established, the atria (or in the case of ventricular fibrillation, the ventricles) are separated from one another by crushing between them the rapid beating persists with little change in each of the separated parts Scherf (1928), as a result of his failure to abolish atrial fibrillation by clamping or ligaturing the atrial muscle in the path of the supposedly circulating waves, was one of the first to throw doubt upon the existence of a circus movement A little later Andrus et al (1930) concluded, from their experimental results in which fibrillation was induced by a single induction shock applied to the auricular appendix, that the waves did not originate and travel in a ring of muscle

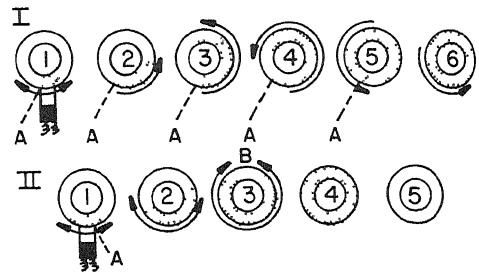


Fig 3 81 (Modified from Lewis 1925) Black = contraction wave Description in text

Evidence against the circus movement theory has been secured by Prinzmetal and his associates (1952) They induced atrial fibrillation by the application of aconitine to the auricles of dogs and observed the auricle by means of high speed cinematography They saw no indication of a circus movement Two types of activity were observed, *a*) contractions and relaxations of microscopic or minute segments of muscle, which they term the "*M*" *activity*, and *b*) stronger contractions and relaxations of macroscopic or larger muscle segments which appear upon the background of the innumerable "M" contractions This type of movement they have designated the "*L*" *activity* Both types of activity are heterorhythmic and utterly chaotic, and apparently arise from numerous discrete ectopic foci

Moe has proposed a variant of the circus movement theory in which he suggests that fibrillation is maintained by the irregular wandering of numerous wavelets generated by the fractionation of a wave front passing through tissue in a state of inhomogeneity with respect to excitability and conduction velocity. He has

devised a most ingenious computer model of atrial fibrillation which supports his theory (Moe et al , 1964)

Direct Current Cardioconversion In the years between 1962 and 1964 the method of direct current cardioconversion devised by Lown (1964) came into wide clinical usage for the conversion of superventricular arrhythmias as well as ventricular tachycardia back to normal rhythm as the apparatus for carrying out this procedure became commercially available The apparatus is designed to deliver a capacitor discharge shock of 100 to 400 watt sec through the chest wall via electrodes 4 in in diameter The shock is timed to fire at the peak of the R wave when conversion of atrial fibrillation, flutter or tachycardia is being carried out, so that the ventricle will be completely refractory when the current hits the heart The atria are depolarized completely, ectopic foci are extinguished, and the opportunity for the sinus node to resume pacemaker function is made available

Lown (1964) has had approximately 95% success in converting superventricular arrhythmias to normal sinus rhythm Maintenance of the sinus rhythm often requires quinidine therapy and even with maximal tolerable doses of this drug, relapse into the arrhythmia may occur, particularly if serious heart disease is present

The cardioconversion apparatus is also effective in abolishing ventricular tachycardia and ventricular fibrillation by administering direct current shock to the heart through the intact chest wall

Ventricular Fibrillation The ventricular muscle may pass into a state of rapid, tremulous and ineffectual contractions closely similar in nature to the condition just described as occurring in the auricle In animals, ventricular fibrillation may be initiated by direct electrical stimulation of the ventricular muscle Mechanical stimulation of the ventricle, especially by pricking the tissue in the AV groove, ligation of a coronary artery or certain chemicals and drugs in excess, e g , digitalis or calcium chloride, may induce fibrillation Chloroform anesthesia renders the hearts of experimental animals (cats) highly susceptible to fibrillation A mere touch of a finger or instrument, the stimulation of a sensory nerve, section of the vagi or their paralysis by atropine, may set the ventricle fibrillating The heart behaves as though sensitized by the anesthetic and ready at the least provocation to fibrillate Adrenaline was found to greatly enhance the effect of chloroform—a fact which indicates the danger of adrenaline administration while a subject is under the effect of this anesthetic Fibrillation may also ensue spontaneously under chloroform, and the condition is more likely to supervene when the animal is passing from deep to light anesthesia Cyclopropane, like chloroform, has the effect of sensitizing the heart to the action of adrenaline in inducing ventricular tachycardia and fibrillation, but these cardiac irregularities are rarely seen under ether anesthesia Quinidine reduces the susceptibility to fibrillation during cyclopropane anesthesia

The effects of ventricular fibrillation upon the circulation are incomparably more serious than those of the corresponding atrial condition This is evident when the importance of the two musculatures in the dynamics of the circulation are compared In fibrillation of the lower chamber the propulsive force of its contraction is practically abolished and the circulation comes to an end, death follows within a few minutes

The basic knowledge of countershock defibrillation acquired through animal experimentation has been applied countless times for the saving of human life The first reports originated from operating room incidents where the exposed heart could be defibrillated by application of the electrodes directly to the epicardium Direct hand pumping massage of the heart was used to sustain circulation until the defibrillating apparatus could be brought to bear Subsequently numerous reports of emergency thoracotomies with open chest massage and defibrillation both in and outside of hospitals appeared Some of these cases were saved, but a distressing number of unsuccessful terminal thoracotomies were carried out

A great advance was made in 1960 when Kouwenhoven, Jude, and Knickerbocker reported the method of closed-chest massage in which circulation is maintained by rhythmic compression of the fibrillating heart through the intact chest wall by strong intermittent manual pressure over the sternum Life can be sustained for periods of time up to an hour or more until external defibrillation can be achieved either by AC or DC countershock through the intact chest

It has been well established that unless circulation can be restored in less than 4 min, irreparable cerebral damage will have occurred, and even though heart action might be restored, a decerebrate patient will be the result

ALTERNATION OF THE HEART

PULSUS ALTERNANS

This is a condition in which every second wave in a pulse tracing is of relatively small amplitude This peculiarity of the arterial pulse is due to alternate variations in the strength of the ventricular systoles, and to a smaller quantity of blood being ejected into the aorta during the weaker beat Figure 3 82 shows a typical sphygmogram of this condition There is as a rule little or no difference in the lengths of the intervals between pulse beats When a slight difference does exist, the interval succeeding the strong beat is then longer than that following the weak beat It will be remembered that in the bigeminal pulse, which might in some instances be confused with alternation, there is inequality in cycle lengths, but the longer interval follows the *weak* (premature) beat Furthermore, in alternation the ventricular rhythm does not share, or does so very rarely, in any irregularity of the pulse intervals which may occur in the arterial tracing The electrocardiogram, for instance, shows no discrepancies in the length of the intervals between the R waves The slight variations in the pulse intervals are attributed by Lewis to a slower rate of transmission of the weaker pulses to the periphery The appearance of alternation in the electrocardiogram (e g , alternate variations in the height of the R wave) is rare When this so-called *electrical alternation* does occur the larger deflections sometimes correspond to the weaker pulse beats But, it will be recalled in this connection that the deflections of the electrocardiogram are determined by the balance of the electrical forces developed during the cardiac contraction rather than upon their total value Electrical alternation bears no relation to the strength of the heart beat



Fig 3 82 Radial tracing showing pulsus alternans (After Mackenzie 1913)

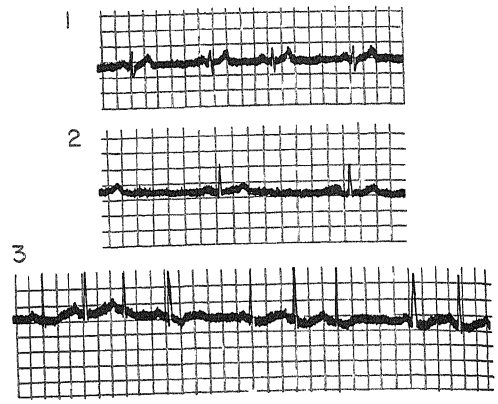


Fig 3 83 / sinus arrhythmia // sinus bradycardia /// sinoatrial block (Kindness of Dr John Hepburn)

Sometimes it is not possible to detect pulsus alternans by palpation of the pulse, the variations in strength of the pulse beat being too slight to be perceptible, but it is clearly revealed in the sphygmogram (Fig 3 82) It may also be detected by means of a blood pressure armlet The pressure in the armlet is raised gradually, when it is found that at a certain level the weaker beats are suppressed, but the stronger beats get through The pulse at the wrist is then precisely half the ordinary rate The pressure during the weaker beats may be as much as 25 mm Hg below that during the stronger, but usually the pressure difference is not more than 5 or 10 mm Persistent alternation of the heart when the pulse is slow or of normal frequency is usually indicative of grave disease of the myocardium Alternation sometimes also occurs at rapid rates of beating, e g , auricular fibrillation, paroxysmal tachycardia etc , but it is then of less serious significance

IRREGULARITIES DUE TO VARIATIONS IN VAGAL TONE

Sinus Arrhythmia (Fig. 3.83I). This is a condition in which rhythmical variations in the rate of the whole heart occur synchronously with respiration It is due to alterations in the strength of the vagal influence upon the pacemaker (SA node) as a result of the respiratory excursions, the heart rate increasing toward the end of inspiration and slowing toward the end of expiration It is an irregularity very com-

mon in children, normal in resting dogs, and may be considered a physiological phenomenon That it is entirely of vagal origin is shown by the fact that it is abolished by atropine It also disappears when the heart rate increases as a result of exercise, fever, etc , but is enhanced by deep breathing

Phasic Irregularity. In this disorder periodic slowing of the heart occurs for a few seconds quite independently of the respirations It also is a vagal effect since it is abolished by atropine The manner of its production is unknown It occurs in convalescence from acute fevers, and sometimes during the administration of digitalis

Sinus Bradycardia (Fig 3.83II). This is a persistent slowing of the whole heart due to increased vagal tone influencing the sinoauricular node The rate may be as slow as 40/min Bradycardia of this nature occurs in apparently healthy persons, many of whom are athletes

Sinoatrial Block (Fig. 3.83III). The entire heart (atria and ventricles) misses a beat at regular or irregular intervals The condition thus differs from AV block in which only the ventricle misses A complete set of waves is therefore dropped from the venous or electrocardiographic tracing, and the arterial pulse intermits Since the condition is temporarily abolished by atropine and may be induced by stimulation of the vagus, it is probable that the missed beats are due to the action of the nerve upon the SA node Sinoatrial block sometimes results from digitalis administration

Atrioventricular Block. Defective conduction between atria and ventricles due to heightened vagal tone is occasionally seen, it is abolished by atropine

4 / MECHANICAL ACTIVITY OF THE HEART

In recent years, the mechanics of cardiac contraction have been examined in accordance with the concepts and methodology developed by A V Hill (1939) and associates for skeletal muscle Isolated strips of myocardium are employed Mechanical activity is analyzed through correlating the interrelations between fiber length, force of contraction, velocity of shortening, tension developed, and load imposed Concepts derived from these in vitro studies are applicable to the analysis of the functional behavior of the intact heart

Muscle Mechanics

The heart is primarily a muscle, which functions in the intact animal as a hydraulic pump Thus, it is appropriate to analyze first the mechanics of cardiac muscle contraction and then apply these results to the functional characteristics of the heart as a hemodynamic pump

THREE COMPONENT MODEL OF MUSCLE

The three-component model of skeletal muscle proposed by Hill (1939) to explain the mechanical properties of skeletal muscle can serve to describe the physiology of contraction of isolated strips of cardiac muscle The three components are (Fig 3 84) *1*) a contractile element (CE) which develops force and shortens when activated, but which is extensible at rest, *2*) a series elastic (SE) element which is in series with the CE and is passively stretched by shortening of CE or stretching of the muscle *3*) A parallel elastic (PE) element which, in the version of the model in Figure 3 84, is depicted as being parallel to both CE and SE It maintains the resting length of the freely extensible CE and supports SE during resting tension, but plays little role during contraction. The anatomical sites of SE and PE are uncertain but may be in the contractile filaments, in the cell membranes, or in both.

In Figure 3 85, the arrangement employed by Sonnenblick et al (1970), for studying muscle mechanics in the isolated papillary muscle from the right ventricle of the cat is diagrammed One end of the muscle is attached to a tension (force) transducer and the other end to an isotonic lever, so that shortening and tension may be recorded simultaneously

Preload is the term applied to the amount of weight attached to the lever which establishes the resting length and tension in the inactive muscles before a contraction is induced *Preload* is analogous to factors which determine muscle fiber length in the intact ventricle at the end of diastole, i e , end-diastolic pressure The lever system is then fixed by means of a stop so as to prevent any further stretching of the muscle *Afterload* is the term applied to any additional weight added after the stop has been fixed, preventing further stretching of the resting muscle This is the additional load which the muscle engages when it begins to contract *Preload* and *afterload* make up the *total weight* against which the muscle is contracting when stimulated The *afterload* component is approximated by aortic pressure in the intact heart

The sequence of events during contraction is diagrammed in Figure 3 86 In the schema the parallel element (PE) is omitted When the muscle is stimulated and excitation-contraction coupling occurs, the contractile element (CE) shortens and develops force This is called the *active state* of the CE (see also p 3-39) The active state represents the time course of the onset and termination of the biochemical processes occurring at active sites along the actin and myosin myofilaments which produce force and shortening Because of the spring-like properties of SE, the active state of CE is translated into mechanical force development or shortening only after some delay The course of mechanical contraction which can be recorded is thus governed by the contractile prop-

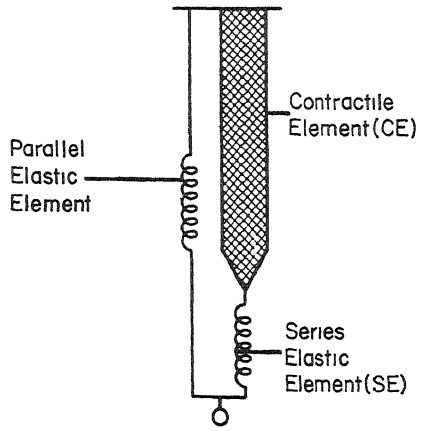


Fig 3 84 Three-component model for muscle proposed by A V Hill (1939) (Redrawn from Sonnenblick 1966)

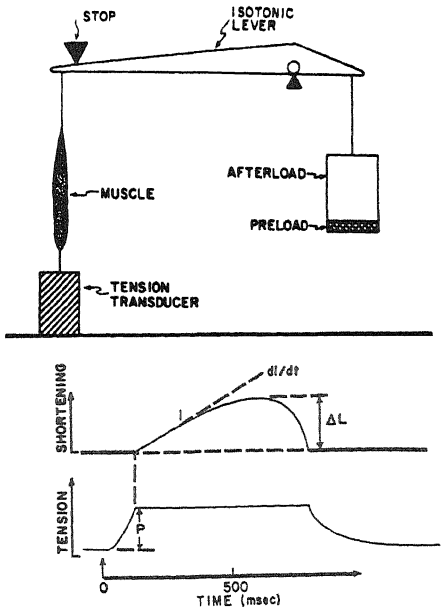


Fig 3 85 *Upper* Arrangement for study of mechanics in isolated papillary muscle *Lower* Records of shortening and tension of afterloaded isotonic contraction P afterload dl/dt initial shortening velocity (Redrawn from Sonnenblick et al 1970)

erties of CE, the duration of CE active state and the elastic properties of SE As seen in Figure 3 86 (*upper*), during an isometric contraction when the muscle cannot shorten, force develops along a delayed time course relative to CE contraction, owing to the time required for stretching SE In Figure 3 86 (*lower*), the model illustrates an afterloaded isotonic con-

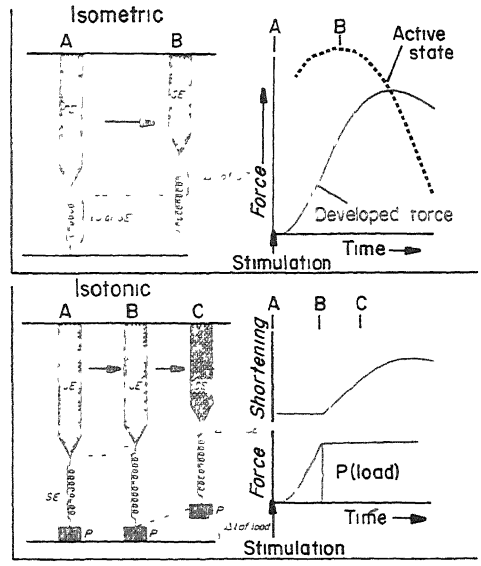


Fig 3 86 Sequence of events during contraction *Upper* isometric contraction showing events in the Hill model (*left*) and curves of force development (solid line) and active state (dashed line) *A* initial resting state *B* an instant during active contraction The difference between the active state curve and developed force curve is owing to the time needed for CE shortening to stretch SE *Lower* afterloaded isotonic contraction (*left*) and curves of force and shortening (*right*) as functions of time (Redrawn from Sonnenblick 1966)

traction The load (*P*) rests on a surface and is not exerting tension (afterload) With stimulation, CE starts to contract and shortens, thus stretching SE When contractile force stretches SE so that its elastic tension equals load (*P*), the muscle begins to shorten and lifts the load (Fig 3 86, *lower*, *B*, *C*)

FORCE-VELOCITY CURVE

During contraction, the force stretching SE remains constant and the course of shortening now represents the shortening of CE In an isolated muscle allowed to contract isotonically against various total loads (preload and afterload), the relation between force and velocity may be determined (Fig 3 87) When preload, which determines initial muscle length, is kept constant, an inverse relation between force and velocity is found The beginning or initial velocity of shortening is plotted as a function of load (*P*) The symbol P_0 is used for the load which just prevents external muscle shortening and thus represents maximum isometric force By extrapolating the force-velocity (or load-velocity) curve to zero load, the theoretical max-

imum velocity of shortening (V_{max}) is obtained The force-velocity relationship of cardiac muscle is changed characteristically by alterations in initial muscle length (preload) and by changes in the fundamental activity level of the shortening processes (contractility, contractile state, inotropic state) When initial muscle length is increased, there is a change in force developed (P, P_0) without a change in extrapolated maximal velocity of shortening (V_{max})[1] (Fig 3 88) In the intact heart this has its counterpart in the Frank-Starling relationship in which the stroke volume and force of contraction can be shown to increase with increasing ventricular volume

Inotropic agents (e g , norepinephrine, calcium, digitalis, glycosides) and procedures increasing frequency of contraction (e g , postextrasystolic potentiation by paired electric stimulation) produce an increase in the V_{max} and P_0 Thus, the muscle shortening is faster for a given load and stronger at isometric loads (see Fig 3 89) Since V_{max} is unchanged by alterations in initial muscle length, but is varied by changes in the contractile state, it can be used as an indicator of the fundamental condition of contractility

The effect of initial muscle length on force of contraction is related to the length of the sarcomeres (Fig 3 90) Alterations of muscle length change the degree of overlapping of the contractile filaments This seems to alter the number of active contractile sites which generate force but not the maximal rate of reaction at these sites This would account for the changes in P_0 with increased muscle length occurring without changes in the V_{max} On the other hand, a change in contractile state appears to result from an influence on the rate of reaction at contractile sites, but not necessarily an effect on the number of sites which are active

[1] There is not general agreement that this relationship holds Although it has been shown that whole cardiac muscle V_{max} is independent of fiber length under specific experimental conditions (i e , afterloaded isotonic contraction at 23° C), there is evidence that this independence is absent under other experimental circumstances Thus, this may not be a general property of cardiac muscle If so, this would limit the value of V_{max} as an index of contractility in cardiac muscle (Pollack, 1970, 1971) However, the concept as described here has been and is widely used as an index of the inotropic state of cardiac muscle and has been vigorously defended (Sonnenblick, 1970) although the controversy remains unresolved

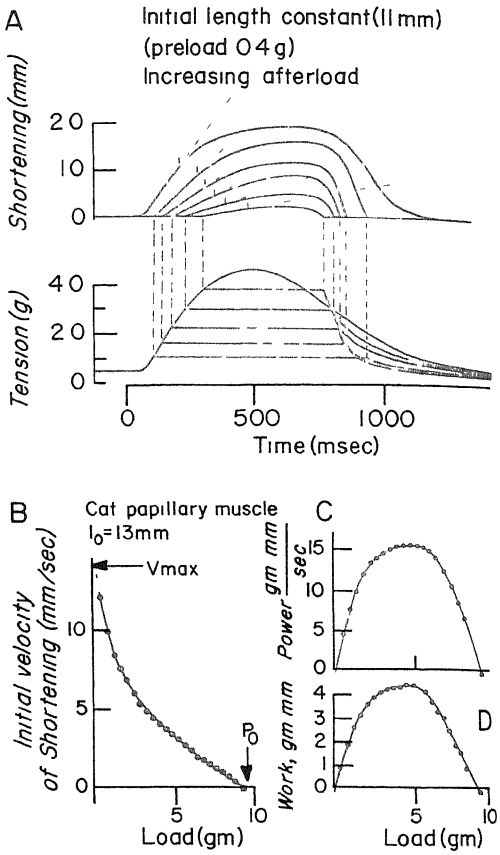


Fig 3 87 *A* curves of the time course of shortening and tension with increasing afterload The initial velocity of shortening (represented by the slope of the dashed lines in the top curves) decreases with increasing afterloads *B* force velocity curve (cat papillary muscle) showing this inverse relation between afterload and initial velocity of shortening *C*, the relation between power $\left(\text{load} \times \text{velocity} = p \frac{dl}{dt}\right)$ and load (p) *D* the relation between work (load $\times$ distance $= \int_{t_0}^{t^1} p \, dl = p \, \Delta L$) and load (Redrawn from Sonnenblick 1966)

LENGTH-TENSION DIAGRAM

As can be seen in Figure 3 91, the force of contraction depends on initial muscle length in both cardiac and skeletal muscle Since muscle length and sarcomere length are directly proportional, the changes in force of contraction with degree of stretch can be related to sarcomere dimensions The extent of thick and thin filament overlap in skeletal muscle determines tension development (Fig 3 92) The evidence supporting these relationships is best for skeletal muscle

The length-tension diagrams are similar in

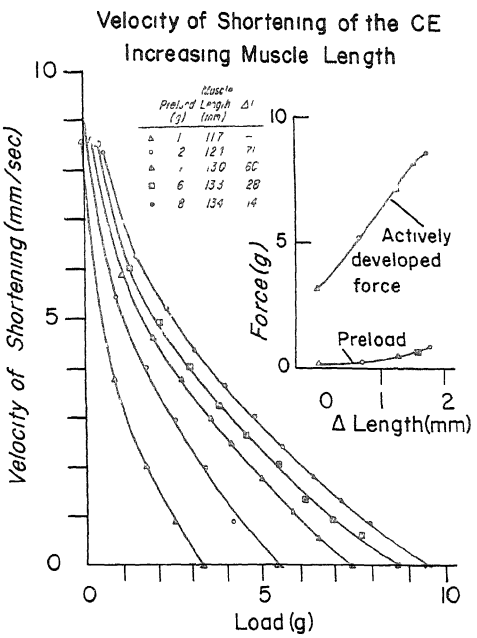


Fig 3 88 Force-velocity curves of cat papillary muscle with increasing preload which increases initial fiber length (insert represents length-tension curve at these preloads) Increasing preload (fiber length) does not alter V_{max} but increases actively developed force (Redrawn from Sonnenblick et al 1970)

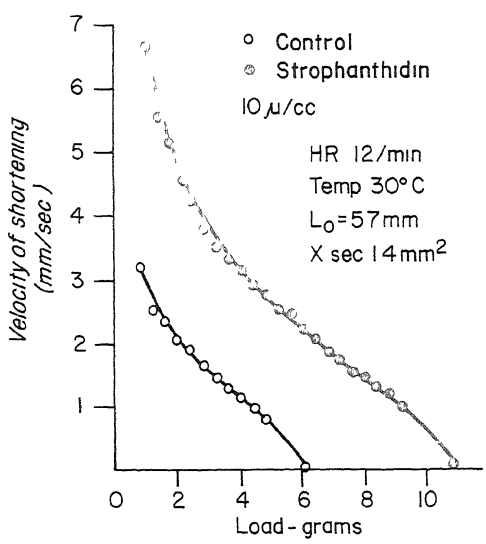


Fig 3 89 Strophanthidin effect on force-velocity relation Both V_{max} and maximal force of contraction are increased (Redrawn from Sonnenblick et al 1970)

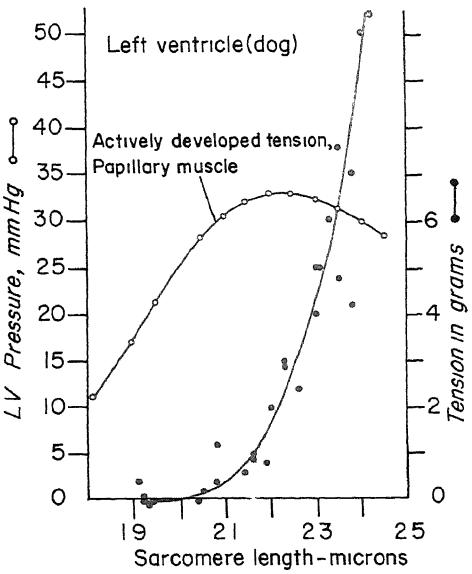


Fig 3 90 Sarcomere length plotted against left ventricular filling pressure (data from dog heart) and sarcomere length-tension diagram from cat papillary muscle Curves are displayed on the same coordinates to indicate the probable relation between sarcomere length and actively developed tension (Redrawn from Spotnitz et al 1966)

form for both cardiac and skeletal muscle However, the maximum force of contraction per unit mass of heart muscle is less than that of skeletal muscle Also, the resting tension of cardiac muscle is considerably higher than that of skeletal muscle at lengths which produce maximum contraction The lesser contractile strength of heart muscle is explained in part because there is a relatively larger amount of noncontractile mass (mitochondria, etc) per cross-sectional area in cardiac than in skeletal muscle When this factor is corrected for and heart muscle is induced to contract at a maximum, the force generating capacity of the two types of striated muscle is similar

The tension curves rise to a maximum and then decline as muscle length is increased (Fig 3 91) The point at which maximum tension develops is termed L_{max} The ascending limb of the length-tension curve represents increasing tension response to increasing length up to L_{max} The descending limb is the part of the curve showing decrease in tension as the initial muscle length is increased beyond L_{max}

There is considerable difference in the resting tension curves between the two muscles

In skeletal muscle, resting tension is very small at L_{max} and only increases significantly during the descending limb of the tension curve In contrast, resting tension in heart muscle is high at L_{max} The structures responsible for this dif-

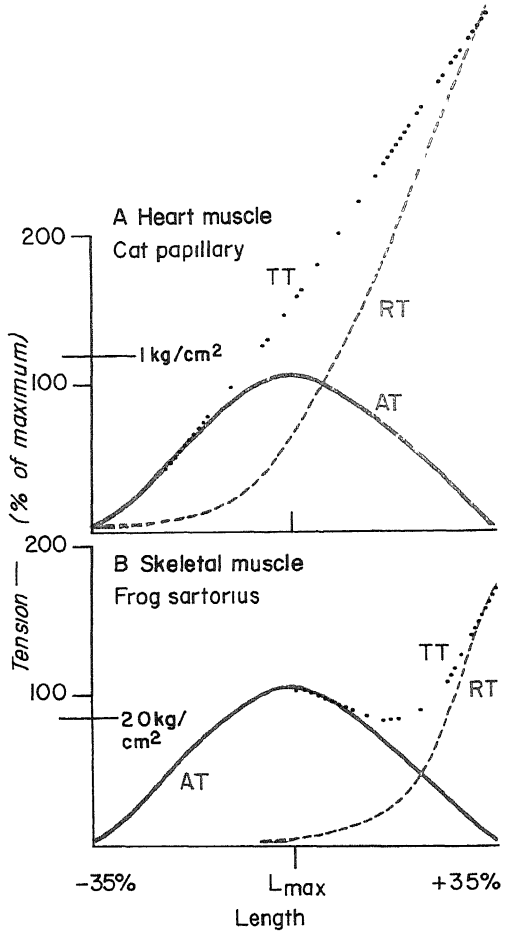


Fig 3 91 Length-tension diagrams for heart (*A*) and skeletal (*B*) muscle compared RT resting tension AT actively developed tension TT total tension (AT + RT) (Redrawn from Spiro and Sonnenblick 1964)

ference in resting tension at L_{max} are not known The connective tissue, sarcolemma, or some other structure may be responsible For cardiac muscle, however, this is a useful feature since the ventricle must function along the ascending limb of the length-tension curve for increasing strength of contraction to be directly related to increases in ventricular volume at the time systole is initiated The cause of this greater stiffness of cardiac muscle may be related to the relatively greater amount of connective tissue present Since cardiac muscle cells are smaller in diameter (5 to 10 μ) than skeletal muscle cells (50 to 100 μ), the relative amount of sarcolemma is greater in the former The sarcomere length-tension curve for papillary muscle is shown in Figure 3 93 The tension curve reaches its apex at sarcomere length between 2 10 and 2 25 μ and tension is close to zero at 1 5 μ As muscle length increases beyond L_{max}, tension decreases but in the descending limb of the curve for cardiac muscle, the relationship between sarcomere length and length of muscle no longer holds strictly and the resting tension increases markedly Thus, sarcomeres longer than 2 60 μ are seldom produced Further, lengthening in response to stretch cannot be due to sarcomere lengthening and must be the result of fiber slippage as explained in the foregoing (see Chap 2, "Dilatation") The effect of sarcomere length on tension development is explained by the "sliding

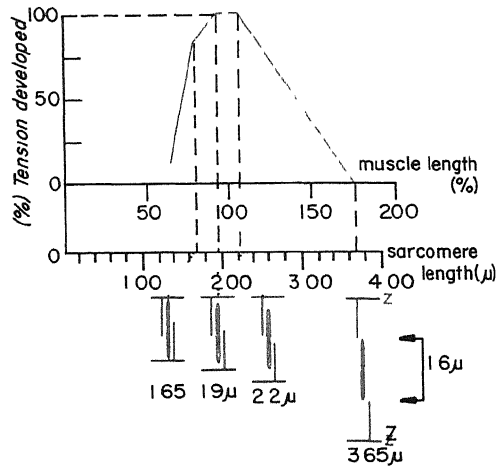


Fig 3 92 Sarcomere length in relation to tension development for skeletal muscle (Redrawn from Hanson and Lowy 1965)

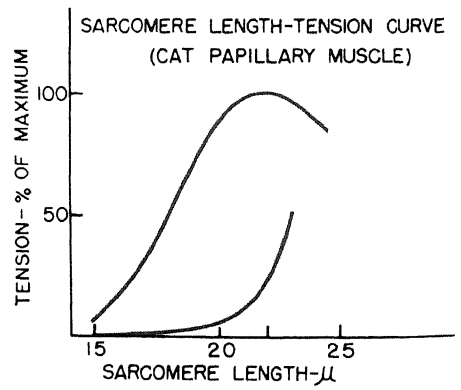


Fig 3 93 Sarcomere length-tension diagram for cat papillary muscle (Redrawn from Sonnenblick, 1968)

filament" theory of muscle contraction (Fig 3 92) Maximal tension depends on optimal overlap of actin and myosin filaments such that all reactive sites between the two contractile protein filaments may be engaged When the sarcomere is stretched, overlapping is decreased and reduction in force development follows If stretched sufficiently, so that no overlap occurs, the muscle becomes incapable of contraction As muscle length is decreased below the apex of the length-tension curve, the actin filaments slide into the center of the sarcomere, this is accompanied by a reduction in force development The latter may be the result of disruption of bond formation in the opposite half of the sarcomere by actin filaments or actual repelling forces developing between them

In the intact ventricle, the volume of ventricular filling determines the initial length of the muscle fibers (preload) At the time of systole the fibers contract isometrically (do not shorten) until the pressure in the ventricle exceeds that in the arteries and ejection begins At this time, shortening of the fibers occurs as contraction proceeds against the pressure in the aorta and pulmonary arteries (afterload) In Figures 3 94 and 3 95, sarcomere lengths, as determined from studies on isolated cat papillary muscles, are plotted to show the relationship

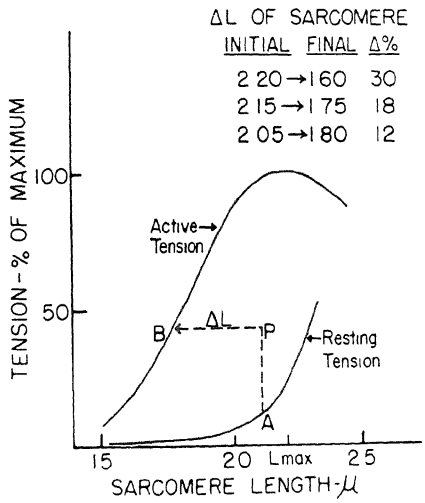


Fig 3 94 The relationship between sarcomere length and active tension during isometric contraction and contraction as the muscle shortens from studies on isolated cat papillary muscle *A* onset of contraction *P* afterload ΔL change in sarcomere length (Redrawn from Sonnenblick, 1968)

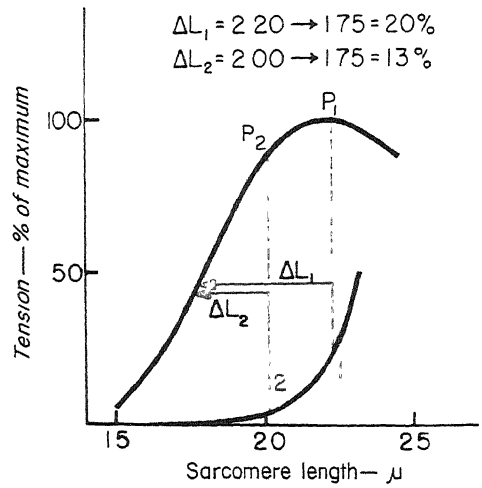


Fig 3 95 The effects of changing initial sarcomere length from 2 to 1 on force of isometric contraction (P_2 to P_1) and extent of shortening (ΔL_2 to ΔL_1) There is little change in contractile force but a substantial change in extent of shortening from 13% to 20% (Redrawn from Sonnenblick 1968)

between sarcomere length and the active tension curve during isometric contraction and contraction as the muscle shortens The latter, auxotonic contraction, is not termed isotonic because the pressure against which the sarcomeres are shortening is changing during ejection (see p 3-96) Force is developed without shortening (Fig 3 94, *A* to *P*) until tension equals blood pressure and the sarcomere shortens (ΔL) The sarcomere then shortens until it reaches the active tension curve (*B*) at which time it can shorten no further In isolated preparations sarcomere shortening of about 30% (from 2 20 to 1 60 μ) may occur (see Fig 3 94). However, under the loading conditions of the intact heart, the sarcomere can only shorten about 12 to 18% In Figure 3 95, the effects of changing initial sarcomere length from 2 00 to 2 20 μ is depicted This is the sole change which would occur in the intact heart with increased filling pressure of the ventricle Because the ventricle operates normally, on the ascending limb of the length-tension curve near its apex (i e , at optimal sarcomere length for developed tension) the resulting increase in force of contraction is only moderate (from P_2 to P_1). However, note that a substantial increase in shortening (ΔL_2 to ΔL_1, 13 to 20%) occurs These fundamental characteristics of cardiac muscle mechanics form the basis of

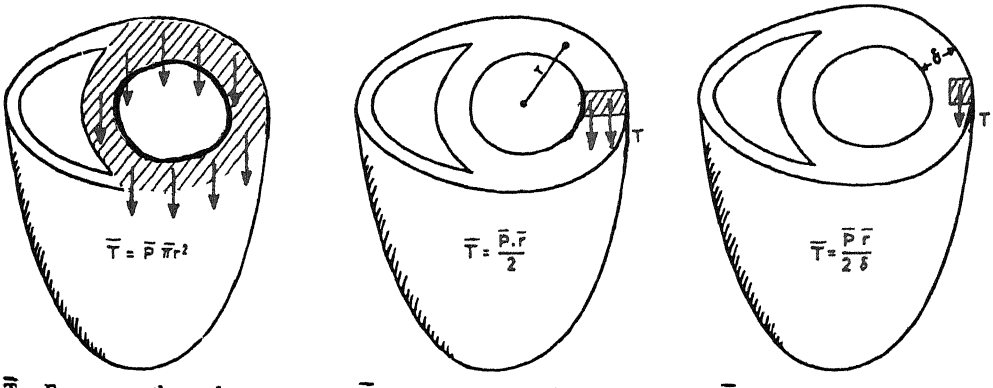
the Frank-Starling relation (Starling's "law" of the heart), according to which, force of ventricular contraction and volume of ejected blood increase directly with an increase in initial volume of the ventricle at the time of systole (end diastolic volume)

VENTRICULAR WALL TENSION AND THE LAPLACE RELATIONSHIP

Ventricular myocardial tension is related to intraventricular pressure per square centimeter of surface upon which the pressure is exerted The larger the surface the greater will be the total pressure This relationship may be calculated by employing the law of Laplace It applies to distensible membranes with a spherical or cylindrical shape Assuming that the ventricle has a spherical cavity, three definitions of contractile tension have been employed for the ventricular myocardium (Fig 3 96) Badeer (1963) recommended expressing myocardial tension in terms of force per unit of cross-section area as follows

$$\bar{T} = (\bar{P}\ \bar{r})/2\delta$$

$\bar{T}$ (stress = mean force per unit cross sectional area of wall during systole, $\bar{P}$ = mean transmural pressure during systole, $\bar{r}$ = mean radius during systole, and δ = thickness of wall during systole Accordingly, a greater pressure or radius (dilatation) increases the contractile tension in the wall of the ventricle (this, of course, applies to the atria as well) The Laplace relationship results in a disadvantage when the ventricle is excessively dilated In the normal ventricle the radius of the chamber during systole decreases as the blood is ejected Thus, the wall tension begins to decrease after the beginning of ejection and at the time of peak systolic pressure may actually be less than at the onset of systole When the ventricle is markedly dilated, the wall tension is greater than normal and the magnitude of fiber shortening may be so compromised that reduction in chamber radius is minimal In such instances, the tension on myocardial fibers may continue to increase from the beginning of ejection to the peak of systolic pressure A further disadvantage of dilatation is that the sarcomeres may be stretched somewhat although there is little anatomical evidence for this Insofar as sarcomeres are stretched, however, the ventricle will be functioning on the descending limb of the length-tension curve Further, the Laplace relation means that, at the same pressure, a dilated heart must generate greater force than normal and, in accordance with the force-velocity relationship, the rate of myocardial fiber shortening will be decreased These factors diminish the ability of the ventricles to eject blood The increased tension in dilated hearts is believed to be adequate stimulus for my ocardial hypertrophy This accounts for hyper trophy in those abnormal states in which the heart is subject to a volume overload at normal



$\bar{T}$ = Force across the total cross-sectional area of muscle

$\bar{T}$ = Force per unit length of circumference and the entire thickness of wall

$\bar{T}$ = Force per unit cross-sectional area of muscle

Fig 3 96 Three ways of representing the relations between ventricular wall tension, intraventricular pressure and ventricular dimensions according to the law of Laplace The ventricle is assumed to be a sphere $\bar{T}$, mean contractile tension (force) $\bar{P}$, mean transmural pressure during systole $\bar{r}$ mean ventricular radius during systole $\bar{\delta}$, ventricular wall thickness (Redrawn from Badeer 1963)

pressure (e g , right ventricular hypertrophy in atrial septal defects)

THE ACTIVE STATE IN THE FORCE-VELOCITY RELATIONSHIP

The concept of the active state has been defined previously (see pages 3-39 and 3-86) The term designates the degree of activity of the contractile element The time course of the active state has been determined on isolated papillary muscles of the rabbit, employing the "quick release" method of Jewell and Wilkie (1960, Edman, 1968) In this method, the muscle is attached to an isotonic lever which, at first, is fixed by a releasable catch so that the muscle begins to contract isometrically At varying intervals during an isometric contraction, the lever is released abruptly and the muscle shortens against a small load The initial rapid phase of shortening involves movements of both the series elastic element and the contractile element A second slower phase is caused by shortening of the contractile element alone The velocity of shortening during the second phase, at a given length change, is an index of the intensity of the active state at that instant of the isometric contraction By releasing the muscle at various times after the stimulus, the time course of the active state may be determined An active state curve of rabbit papillary muscle is shown in Figure 3 97 and the relation of active state to isometric contraction on the right top in Figure 3 86 The active state attains a maximum slowly and starts to decay without reaching a plateau The latter suggests that the contractile element of the myocardium does not reach mechanical "saturation" during contraction It has been shown (Edman, 1968) that the course of the active state depends on how much the contractile element is allowed to shorten during the contraction period As the contractile element is allowed to shorten, there is a progressive decrease of the maximal intensity of the active state and a reduction in its duration Thus, shortening of the muscle during contraction reduces the duration of mechanical activity (see Figs 3 98 and 3 99)

THREE DIMENSIONAL DIAGRAM OF VENTRICULAR CONTRACTION

The force-velocity curve and length-tension curve for the papillary muscle can be arranged

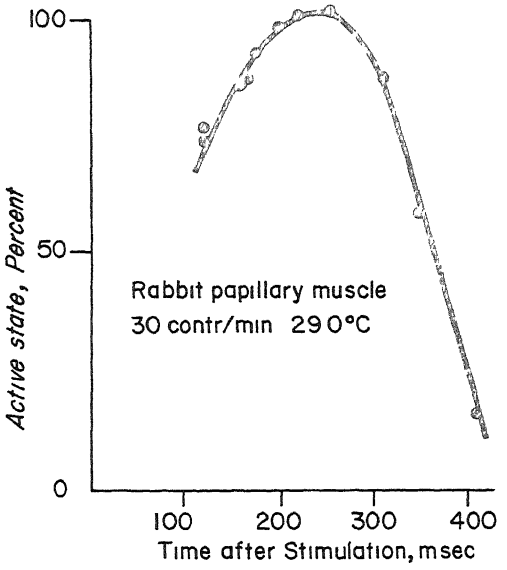


Fig 3 97 Active state curve of rabbit papillary muscle (Redrawn from Edman 1968)

in a three dimensional diagram to illustrate their interrelations, in particular the effect of change in muscle length on the force-velocity curve Such a diagram is seen in Figure 3 100 The vertical plane on the right represents the force-velocity relation, the vertical plane on the left the velocity-length relation, and the triangle at the base the length-tension relation A series of force-velocity curves with decreasing fiber length is displayed The course of the contraction is depicted by the heavy dashed line The muscle is stimulated at point A and, after an appropriate interval for the development of the active state, the CE reaches point B on a given force-velocity curve Note that the force-velocity curve followed depends on the muscle length at which the CE begins to develop force As force (load) increases, the velocity of shortening of the CE travels down the force-velocity curve to some point C At this point C, force of contraction equals load (afterload) and mechanical shortening starts at that load (isotonic contraction) The velocity of shortening now decreases as the length of CE decreases, traveling down the family of instantaneous force velocity curves The active state, and thus, the velocity of shortening of the CE at any instant, depends on the muscle length at that moment in time Thus, during contraction, as the muscle shortens, it operates on a series of instantaneous force-velocity planes of decreasing magnitude

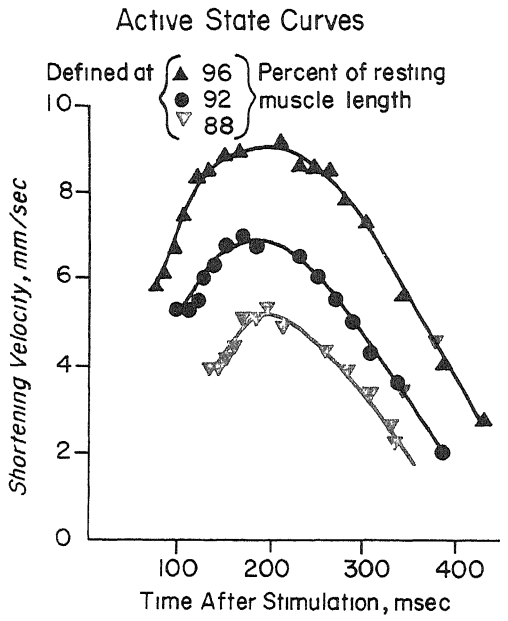


Fig 3 98 Changes in active state curves of papillary muscle in relation to the degree of muscle shortening permitted Shortening of the contractile unit decreases the intensity and duration of the active state (Redrawn from Edman 1968)

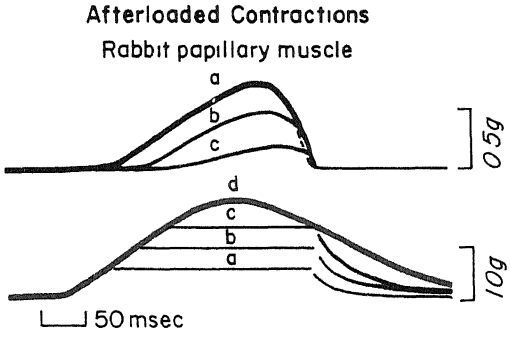


Fig 3 99 Relationship between papillary muscle shortening and duration of mechanical activity Isotonic contraction curves a b c (*upper*) superimposed on the same time base with isometric myogram c b a (*lower*) Curve d is the ordinary isometric contraction myogram The more the muscle is allowed to shorten during the isotonic phase the briefer is the isometric portion of the myogram (Redrawn from Edman 1968)

PRESSURE-VELOCITY-VOLUME (PRESSURE-VELOCITY-LENGTH) DIAGRAM FOR THE INTACT VENTRICLE

In the intact ventricle, pressure, velocity, and volume are analogous to force, velocity, and length of the isolated cardiac muscle (Fig 3 101) Here, force is analogous to developed

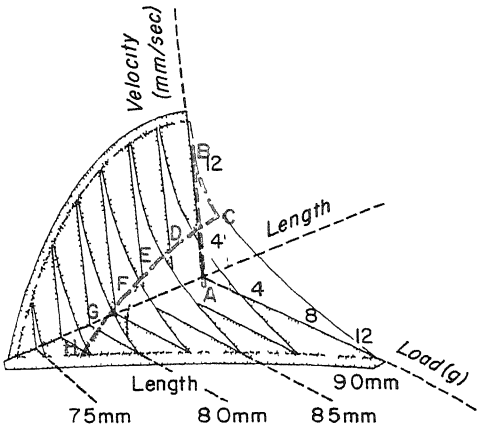


Fig 3 100 Three-dimensional diagram representing the interrelations between force-velocity and length tension curves of papillary muscle Base length-tension curve Vertical plane on right force-velocity curve Vertical plane on left velocity-length curve Dashed line (ABCDEFGH) the course of a single afterloaded contraction (Redrawn from Sonnenblick 1966)

pressure, length analogous to volume, and muscle shortening analogous to the amount of blood ejected (stroke volume) In Figure 3 101, work is illustrated on the horizontal plane as the product of force (or load, obtained from pressure) times distance (muscle shortening or stroke volume change) On the vertical plane, power is represented as the product of force (load) times velocity As in Figure 3 100, the dark line represents the course of a single contraction At A, contraction is initiated and CE velocity rises toward V_{max} as CE begins to shorten It travels down the force velocity curve to point C This represents isometric or isovolumetric contraction of the ventricle At C, force equals load and mechanical shortening begins traveling from C to E (ventricular ejection, stroke volume) The area ADEF represents stroke work (pressure $\times$ volume) The area ABCD represents power (pressure $\times$ velocity) The effect on these relationships of increasing ventricular volume at the onset of contraction is shown on the curve HJICE The maximum isometric force is increased from P_0 to P_0' Little change in V_{max} is seen but both stroke work and stroke power are increased This is the change which occurs with increasing end diastolic volume of the ventricle There is little change in the contractile state of the myocardium since V_{max} is not changed The insert in the lower right shows predicted relations of stroke volume, stroke work, and

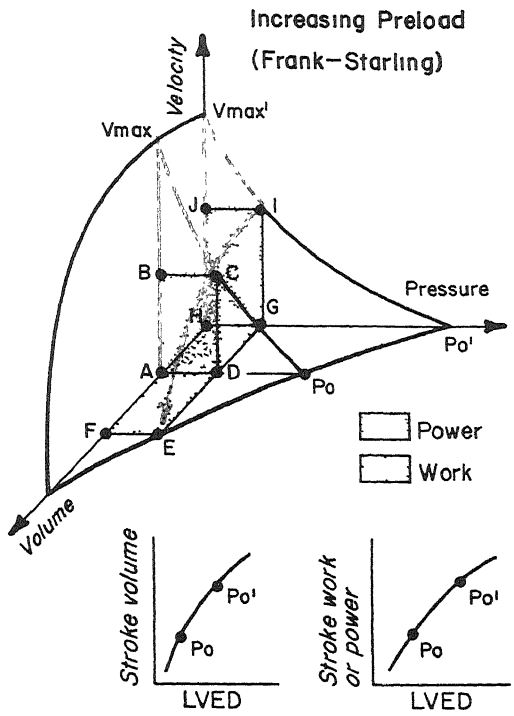


Fig 3 101 The effects of increasing preload on the work and power in analogy to pressure volume relations in the intact ventricle Load is represented as pressure and volume as length on the three-dimensional force-velocity-length construct for papillary muscle When preload (analogous to left ventricular end diastolic pressure LVED) is increased and pressure remains the same work increases (ADEF to HGEF) and power increases (ABCD to HJIG) V_{max} representing contractile state is unchanged *Insert* Relations between stroke volume and stroke work or stroke power predicted from the changes in work and power areas of the three dimensional diagram (Redrawn from Sonnenblick 1966)

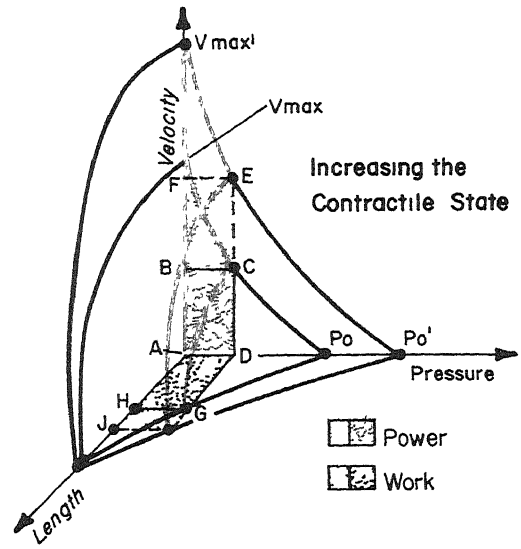


Fig 3 102 The effects of increasing contractile state with inotropic agents (sympathetic nerve stimulation norepinephrine) on force-velocity-volume (length) relations V_{max} and P_0 are increased Load (pressure) is kept constant Work and power are increased (Redrawn from Sonnenblick 1966)

stroke power to left ventricular end diastolic pressure The effect of increasing contractile state (e g , as with increase of sympathetic tone, norepinephrine) is shown in Figure 3 102 Note that an increase in both V_{max} and P_0 occurs and velocity of ejection of the ventricle is increased at any given intraventricular pressure and volume Work and power increase as shown Observe that the pressure (distance from A to D) has not been increased

In Figure 3 103, the effects of increasing heart rate are portrayed An increase in V_{max} without a change in P_0 occurs Power is augmented from ABCD to AFED, while work remains the same, ADGH Ventricular performance curves in Figure 3 103 relate left ventricular end-diastolic pressure to stroke volume, work, and power These show that there is no increase in the first two but the stroke power curve is shifted to the left

Figure 3 104 illustrates the effect of alterations in blood pressure (afterload) alone Both power and work may be increased with increasing aortic pressure from P to P' in the absence of a change in contractile strength The effects of afterload increases on ventricular performance are shown in the insert. With increased afterload the initial velocity of shortening is decreased (V to V') and the extent of shortening is decreased Thus, the ventricular function curve relating stroke volume to left end diastolic pressure (Fig 3 104) is shifted to the right as stroke volume and peak ejection velocity decrease Both stroke work or stroke power may be increased at any given left ventricular end diastolic pressure, thus shifting this function curve to the left (bottom, Fig 3 104) Therefore, the effect of increasing afterload or blood pressure may be limited to mechanical changes alone with no influence on contractile state (however, see Anrep effect, p 3–110)

The three dimensional diagram in Figure 3 105 represents more realistically the relations in the intact ventricle, since part of ventricular systole is *auxotonic*, i e , contraction against an

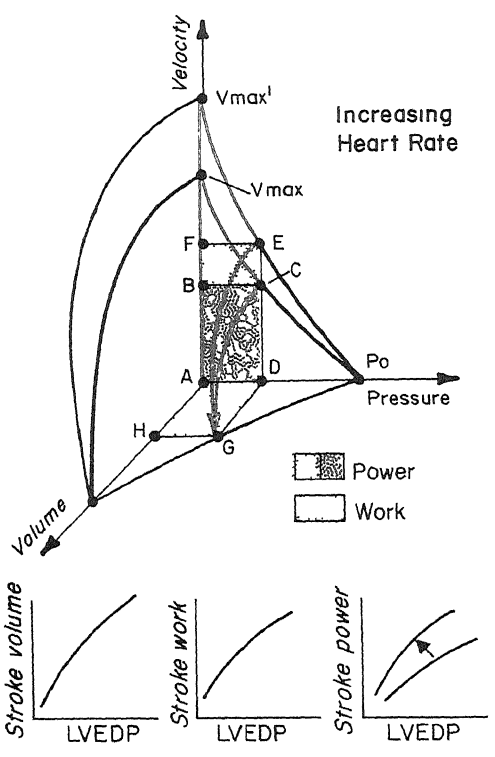


Fig 3 103 The effects of increasing heart rate on force-velocity-length relations V_{max} is increased without a change in P_0 Power is increased with no change in work *Insert* Relations between stroke volume stroke work and left ventricular end diastolic pressure (LVEDP) show no change but the stroke power curve is shifted to the left (Redrawn from Sonnenblick 1966)

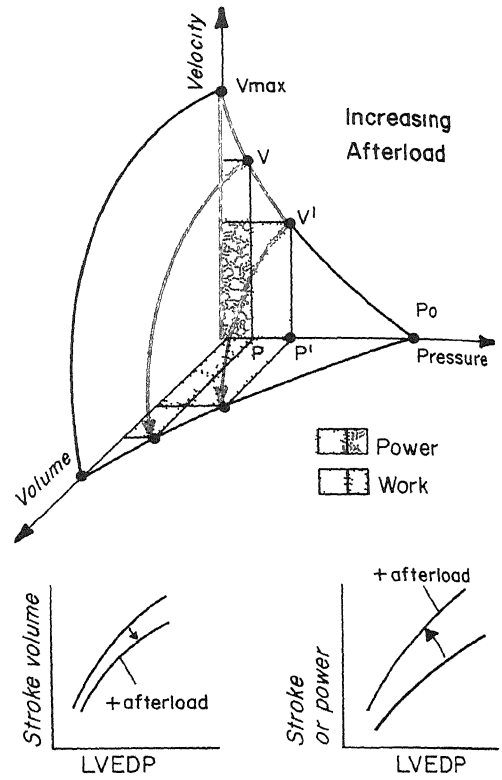


Fig 3 104 The effects of increasing afterload (analogous to blood pressure) on power and work Work and power are both increased with no change in V_{max} (i.e., contractile state remains constant) *Insert* In the intact ventricle with no increase in contractile state elevation of blood pressure would tend to diminish stroke volume relative to left ventricular end diastolic pressure (LVEDP) but stroke work and power would be augmented However in the intact heart contractile state may be augmented with a pressure increase (See Anrep effect p 3-110) (Redrawn from Sonnenblick 1966)

increasing pressure [2] Velocity of CE shortening follows the usual course until the onset of ejection (point B) Fiber shortening is then determined by the varying afterload (blood pressure) and the instantaneous relation between fiber length and velocity Point C marks the end of ejection, the question mark indicates that its exact location relative to the length-tension curve is not certain The length-tension curve depicted is determined from isovolumetric beats in which the ventricle is caused to contract with the aorta clamped or obstructed by a balloon so that ejection cannot occur Following isovolumetric relaxation (C to D), ventricular filling occurs and fiber length increases along the line D to A Point A is end diastole and beginning systole

[2] Actually, these relations are more complex, since during ventricular ejection the pressure first increases and then begins to decline before contraction (systole) is complete and ventricular wall tension relative to pressure (Laplace relation) decreases as ventricular radii decrease during systole The term "negative" auxotonic contraction is sometimes used for contraction against a decreasing pressure Thus, the ejection phase of ventricular contraction can be described as "positive" auxotonic initially and "negative" auxotonic terminally See also p 3-91

WORK, POWER, AND EFFICIENCY

In simplest terms work (W) equals force (P) times distance or displacement (ΔL)

$$W = P \Delta L$$

and power ($\dot{W}$), the rate of work or the work performed per unit time (t)

$$\dot{W} = W/t = P \Delta L/t$$

or force times velocity (dl/dt or v) $\dot{W} = P\ dl/dt = P\ v$

Thus, work performed is the power integrated as a function of time

$$W = \int (P\ dl/dt)\ dt = \int_{t_0}^{t_1} Pdl = P\ \Delta L$$

In Figure 3 87 *C* and *D* are depicted the relations between load and power and work development by the contracting papillary muscle The kinetic component of work equals ½ mv^2, comprises less than 5% of work, and has been omitted

From Figure 3 87 *B*, *C*, *D*, it can be seen that for any given muscle length, work and power vary with load (*P*) reaching a maximum at 50% of the maximum force (P_0) the muscle can develop at that length When shortening does not occur (at P_0 and at the end of contraction, V_{max}) the load is not moved through any distance Distance becomes zero and both work and power reach zero Thus, when mechanical shortening of the muscle does not occur, no work, in a physical sense, is done In another sense, however, work is performed in isometric contraction as represented by shortening of the contractile elements (CE) in the total muscle

In Figure 3 106, various expressions of cardiac work are represented *Contractile element work* (shortening of CE against SE) has been explained in the foregoing *Total cardiac work* includes that produced during the entire contraction period (systole) This would also incorporate work expended in propelling blood in

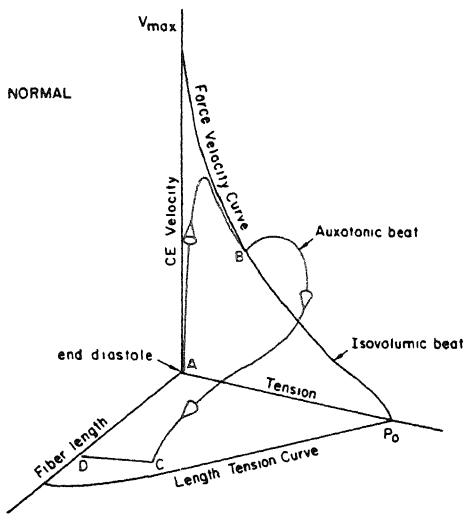


Fig 3 105 Three dimensional construct for the intact ventricle (Redrawn from Braunwald et al 1967)

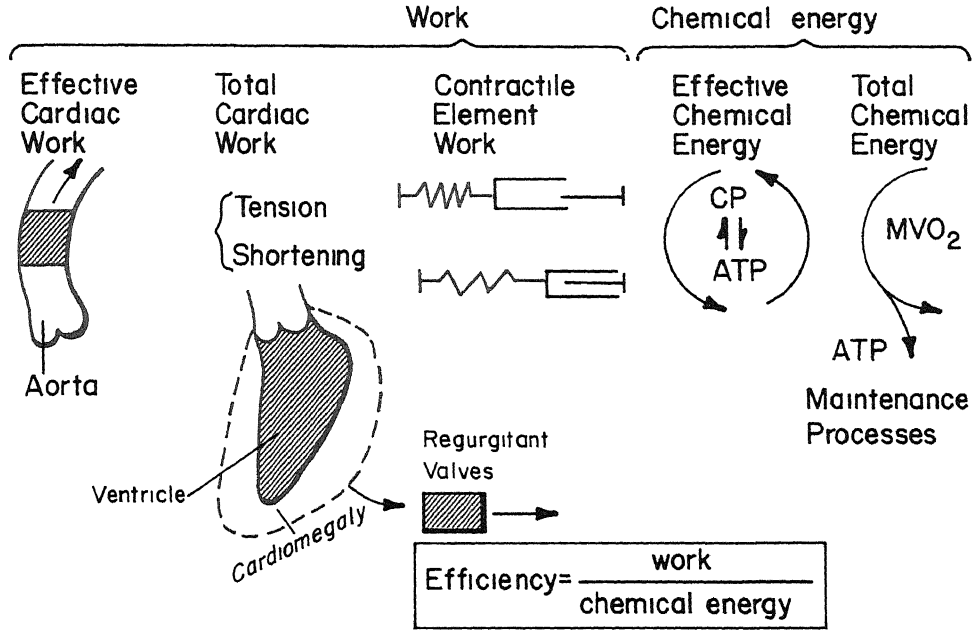


Fig 3 106 Various factors employed to calculate cardiac efficiency Chemical energy can be taken as total myocardial oxygen consumption (MVO_2) or by effective energy used in the contractile process (ATP utilization) Work can be considered as merely contractile element (CE) shortening against load or by the total cardiac work (which is a product of volume of blood ejected against pressure) The latter is affected by ventricular geometry dimensions and regurgitation of cardiac valves (pp 3-92 and 3-44) Effective cardiac work considers only that calculated from the stroke volume delivered forward into the great vessels (e g aorta) against a given pressure (Redrawn from Braunwald et al 1967)

the wrong direction (regurgitation) should cardiac valves be insufficient *Effective cardiac work* is that which produces forward movement of blood against arterial pressure This includes only that work accomplished during systolic ejection when a volume (stroke volume) of blood is delivered into the aorta (or pulmonary artery) against the existing arterial pressure

Efficiency (Eff), most simply defined, is the ratio of useful energy output (work) to total energy expenditure (work + heat)

$$\text{Eff} = \text{work}/(\text{work} + \text{heat}) = w/w + h)$$

Work plus heat is the equivalent of total chemical energy expended (e) Thus, efficiency can be represented as

$$\text{Eff} = W/e$$

as in Figure 3 106

Since practically all energy consumed in the myocardium is linked to oxidative metabolism, myocardial oxygen consumption (MVO_2) can be considered an equivalent of total chemical energy Then the efficiency ratio becomes

$$\text{Eff} = \frac{W(\text{pressure} + \text{kinetic})}{\text{Total } MVO_2 - \text{resting } MVO_2}$$

Values for cardiac efficiency calculated in this way have varied between 10 and 20% in experimental animals This expression of efficiency relates only oxygen consumption to effective cardiac work (Fig 3 106) and neglects a number of factors linking the chemical and mechanical processes Effective ventricular work excludes part of total cardiac work and part of oxygen consumption represents noncontractile processes

A closer approximation to true cardiac efficiency is the ratio of contractile element work (CEW) to chemical energy of activation (AE) and chemical energy of work (WE) less resting energy (RE)

$$\text{Eff} = \text{CEW}/[(\text{AE} + \text{WE}) - \text{RE}]$$

An efficiency of 33% was obtained for the anaerobic iodoacetate-treated papillary muscle using this approach (Pool and Sonnenblick, 1967) When RE was not subtracted the efficiency obtained was 23%

The Heart as a Pump

ATRIAL AND VENTRICULAR BLOOD VOLUME

In man, the atrial capacities are slightly greater than those of the corresponding ventricles, thus providing a reservoir of blood sufficient to fill the ventricular cavities completely For example, in the dead human heart such values as the following have been found the right atrium 163 ml , the right ventricle 137 ml , the left atrium 140 ml , the left ventricle 120 ml

It is now well established that the heart is not completely emptied of blood during ventricular contraction, i e , a residual ventricular volume of blood remains at the end of ejection Observations in dogs indicate that the right ventricle and left ventricle both empty in a fractionate manner

Under control conditions in anesthetized dogs, ejection fraction averaged 47% of left ventricular end diastolic volume With reduction of outflow resistance (afterload), ejection fraction increased to 73%, with little change in stroke volume (Tsakiris et al , 1968) Caution is necessary, however, regarding the methodology of in vivo ventricular volume measurements Two such methods are currently available one is based on the indicator dilution technique, measuring during successive heart beats the decline in concentration of the indicator introduced into the ventricle, the other utilizes cineradiographic measurements following the injection of a radiopaque substance outlining the ventricular cavity Theoretical objections to the former method include the demonstrated incomplete mixing of the indicator within the ventricular chamber, while angiographic methods must necessarily involve certain geometric assumptions, in addition to the unknown effect on the heart of the rapid injection of a relatively large volume of contrast material

In man, a roentgenographic method gave left ventricular diastolic volumes of 70 to 80 ml ± 10 to 20 ml /sq meter of body surface, with 60 to 75% ejected during systole (Dodge et al , 1966) With a thermodilution method diastolic volumes of about 100 ml /M^2 and about 50% of the diastolic volumic ejected during systole were reported (Rapaport, 1966) Stroke volume (SV) is the difference between end-diastolic volume (EDV) and end-systolic volume (ESV) The ratio SV/EDV has been shown to be quite constant in mammals varying in size from rat to horse and to vary linearly with body weight (Holt et al , 1968) In man and dog it varies between about 50 and 65%. Sonnenblick (1968) explains this proportion constancy in hearts of vastly different size on the basis of sarcomere

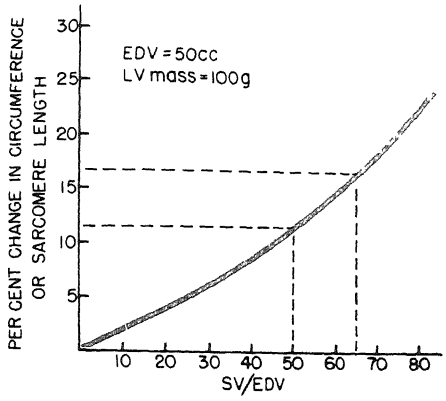


Fig 3 107 The percentage change in midwall circumference or sarcomere length to ventricular ejection fraction (SV/EDV stroke volume/end diastolic volume) (Redrawn from Sonnenblick 1968)

characteristics, which, in all species, have similar dimensions In Figure 3 107, percentage change in ventricular midwall circumference and sarcomere length is plotted against ejection fraction (SV/EDV) The dashed lines represent ejection fractions within the reported range of 50 to 65% This would require sarcomere shortening of 12 to 17%, which corresponds well with that predicted from afterloaded sarcomere shortening Thus, the characteristics of sarcomeres, which are uniform in different-size hearts, would explain ventricular ejection and apply to hearts of any size

PATTERNS OF VENTRICULAR EMPTYING

Traditional concept has it that the ventricles eject blood primarily by shortening of the longitudinal axis accompanied by a rotational action comparable to a wringing out of a wet rag Direct measurements, however, of changing ventricle dimensions in unanesthetized dogs by cinefluorographic techniques with contrast media, by intraventricular gauges and by external resistance gauges, indicate that the left ventricle resembles a cylinder with a conoid apical segment The cylindrical portion is made of a strong cuff of deep fibers between thin layers of spiral muscle The conoid section is made primarily of closely woven spiral muscle entering and leaving "the vortex." Left ventricular contraction involves primarily a reduction in transverse diameter This action accounts for most of the power and volume of the ejection since the contained volume decreases with the square of the radius in a cylinder There is relatively little rotation or shortening of the longitudinal axis All this would be expected since the bulk of the fibers is circularly arranged

By contrast, right ventricular ejection of blood can be effected by three means *1)* Longitudinal shortening of the chamber, i e , base moving toward apex is the most obvious movement This might be expected since the inner and outer layers of spiral muscle making up most of the right ventricle are oriented at about 90° from each other Hence, simultaneous contraction of the two layers of spiral muscle produces shortening along the longitudinal axis of the right ventricle chamber *2)* The right ventricular chamber is roughly triangular in shape, being bounded by a convex septal well and a concave free or lateral wall which enclose a crescent shaped area between them The free wall of the right ventricle moves toward the convex surface of the septum This movement although slight could be extremely effective in moving blood, and it would operate like a bellows used to kindle a fire, i e , since the sides of the ventricle or bellows are large compared to their enclosed space, slight movement toward each other should cause displacement of a large volume of blood *3)* Contraction of the left ventricle must produce a greater curvature of the septum, and thus, since this is attached to the right ventricular lateral wall, traction on it will add to the bellows action The action of this or the other factors has not been quantitated That it can be a potent mechanism, however, is borne out by the observation that right ventricular ejection can be maintained when the free wall of the right ventricle has been almost completely destroyed by cautery in the dog or by coronary occlusion in man

The reduction in diameter and circumference necessary to eject the stroke volume is different in the two ventricles because of their shape and manner of contraction but they are small in both cases Under normal circumstances the ventricles function at relatively large diastolic and systolic dimensions since possibly half their contained volume is not ejected with each contraction If the left ventricle is regarded roughly as a cylinder which ejects approximately half its contained volume of 130 ml , then the diameter can decrease from 5 2 cm to 3 9 cm , and the circumference from 16 5 cm to 12 1 cm

The right ventricle has some semblance to a segment of a sphere, and a small shortening of its fibers will eject a large blood volume If its volume is also 130 ml , and half is ejected, the diameter will decrease from 6 3 cm to 5 1 cm ,

and the circumference from 19 8 cm to 16 0 cm The average shortening of the right and left ventricle walls might be estimated as 20% and 25% of the diastolic size The extent of myocardial shortening also varies greatly for the same volume of ejection depending on the size of the cavity or the fiber length before ejection takes place It takes a much smaller reduction in fiber length of a larger sphere or cylinder to eject the same volume For example, if in a heart enlarged by congestive failure of the circulation the right ventricle contains 500 ml blood, then ejection of 60 ml will decrease the diameter from 9 8 cm to 9 4 cm , or by 0 4 cm , and the circumference from 30 8 cm to 29 5 cm

It would be expected that the shortening of muscle fibers of the lateral wall of the right ventricle at different depths would be similar since the wall is so thin In the left ventricle, however, the circumferentially arranged deep constrictor fibers form a cuff of muscle making up most of the muscle wall which is so thick that the inner layers near the endocardium have a much smaller radius and circumference than the outer layers near the epicardium With the onset of contraction, the inner layers must shorten more than the outer layers in ejecting a particular volume Thus, presumably, no two layers of the myocardial fibers shorten to the same extent during ejection

The following ventricular volumes, somewhat analogous to lung ventilatory volumes, are identified (Fig 3 108) *stroke volume*, the quantity of blood ejected during systole, *residual volume*, the quantity of blood remaining in the heart after a maximal systolic ejection, *systolic reserve volume*, the quantity of blood remaining in the heart at the end of systole, minus the residual volume, *diastolic reserve volume*, is the maximal amount of blood which the heart can, under physiological conditions, receive and then eject, in addition to the normal diastolic inflow Thus, it is the potential volume increase which the heart can accommodate under physiological conditions

Rushmer et al , (1958, 1970) have shown in studies of ventricular dimensions (diameter) and stroke volume in dogs, that the left ventricle functions close to its maximal diameter while the animal is reclining at rest (Fig 3 109), i.e , little diastolic reserve volume is maintained Upon standing, ventricular dimensions and stroke volume diminish Now, with exercise, stroke volume can be increased by increasing diastolic distention (decreasing diastolic reserve) or increasing systolic ejection (decreasing systolic reserve) or both (Fig 3,109) See pages 3-44 and 3-91 for a discussion of the effects of dilatation (increasing diastolic distention) on ventricular contraction Large diasolic size favors ventricular contraction by

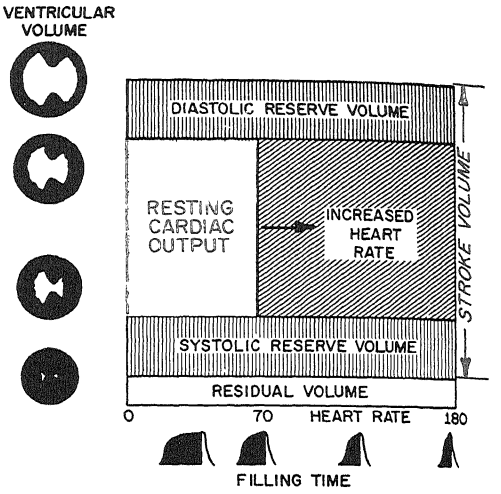


Fig 3 108 Ventricular volumes Resting cardiac output (about 5 liters/min) equals stroke volume (about 70 ml) times heart rate (about 72/min) Maximal cardiac output increase of about sixfold (total shaded area) can occur through elevation of heart rate and stroke volume Stroke volume increases by encroachment on systolic and diastolic reserve volumes (Redrawn from Rushmer 1970)

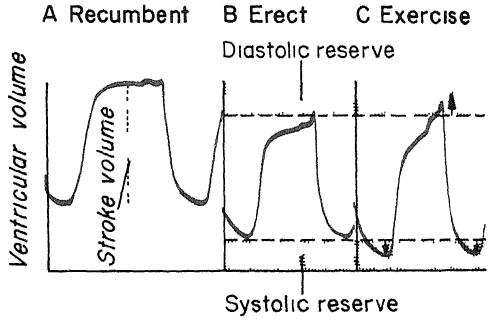


Fig 3 109 Alterations in ventricular volume with postural changes and exertion *A* in supine recumbency diastolic volume is near maximal and stroke volume large *B* upon assuming erect posture diastolic and stroke volumes decrease and the heart has diastolic and systolic reserve volumes *C* with bodily exercise stroke volume increases at the expense of these reserve volumes (Redrawn from Rushmer 1970)

) the Frank-Starling mechanism (pp 3-44, 3-19, and 3-91, 2) the larger volume ejected for i given amount of fiber shortening (see p 3-100), and 3) reduction of internal friction within the myocardium Opposing these favorable factors is the increase in wall tension required to eject blood at a given pressure (Laplace relation, see p 3-92) See page 3-129 for a discussion of cardiac reserve

PRESSURE AND VOLUME EVENTS OF A CARDIAC CYCLE

For describing the events of the cardiac cycle, terminology and symbols according to Wiggers (see Table 3 12) and the schema and time intervals (Fig 3 110) by Selkurt and Bullard (1966) are used

Sequence of Events The upper set of pressure pulse curves in Figure 3 110 are for the left heart and aorta and the lower set for the right heart and pulmonary artery These events are correlated with periods when semilunar and atrioventricular valves are opened and closed, with heart sounds, and with the electrocardiogram (*lower*, Fig 3 110) The phases of the cardiac cycle according to Wiggers are identified by Roman numerals

I Isovolumetric Contraction At the time of the peak of the R wave in the electrocardiogram, ventricular contraction begins The pressure curve starts to ascend and at the moment it exceeds the atrial pressure curve, the atrioventricular valves close The "C" wave in the atrial curves begins at this time, caused by bulging of the closed AV valves into the atria as ventricular pressure increases The "C" waves are variable and may be absent (dotted lines, Fig 3 110) Both atrioventricular and semilunar valves are now closed and the ventricles are contracting on the contained blood which is incompressible Therefore, the volume of the ventricles does not change and pressure increases rapidly This period terminates and the next period begins at the moment ventricular pressures exceed aortic or pulmonic pressures The semilunar valves open and the blood is accelerated into the great arteries

II Maximum Ejection This period begins with the opening of the semilunar valves Most of the blood ejected during systole flows during this period, as can be seen by the rapid descent of the left ventricular volume curve It lasts until the peak of the arterial pressure curve For about one-third of the period of

Table 3 12
Events of the cardiac cycle (Wiggers)

Phase	Interval	Events at Onset of Phase	Main Events during Phase	Events at Close of Phase
Isovolumetric contraction	I	Onset of ventricular contraction	Closure of atrioventricular valves, rapid rise of intraventricular pressure	Opening of semilunar valves
Maximum ejection	II	Opening of semilunar valves	Rapid outflow of blood from ventricles	Peak of intraventricular pressure
Reduced ejection	III	Peak of intraventricular pressure	Declining outflow of blood from ventricles	Onset of ventricular relaxation
Protodiastole	IV	Onset of ventricular relaxation	Rapid drop in intraventricular pressure	Closure of semilunar valves
Isovolumetric relaxation	V	Closure of semilunar valves	Continued ventricular relaxation with no volume change	Opening of atrioventricular valves
Rapid inflow	VI	Opening of atrioventricular valves	Rapid flow of blood from atria to ventricles	Slowing of inflow rate
Diastasis	VII	Slowing of inflow from atria to ventricles	Continued slower flow from atria to ventricles	Onset of atrial contraction
Atrial systole	VIII	Onset of atrial contraction	Increased flow from atria to ventricles	Termination of atrial, and onset of ventricular, contraction

From Dittmer and Grebe, 1963

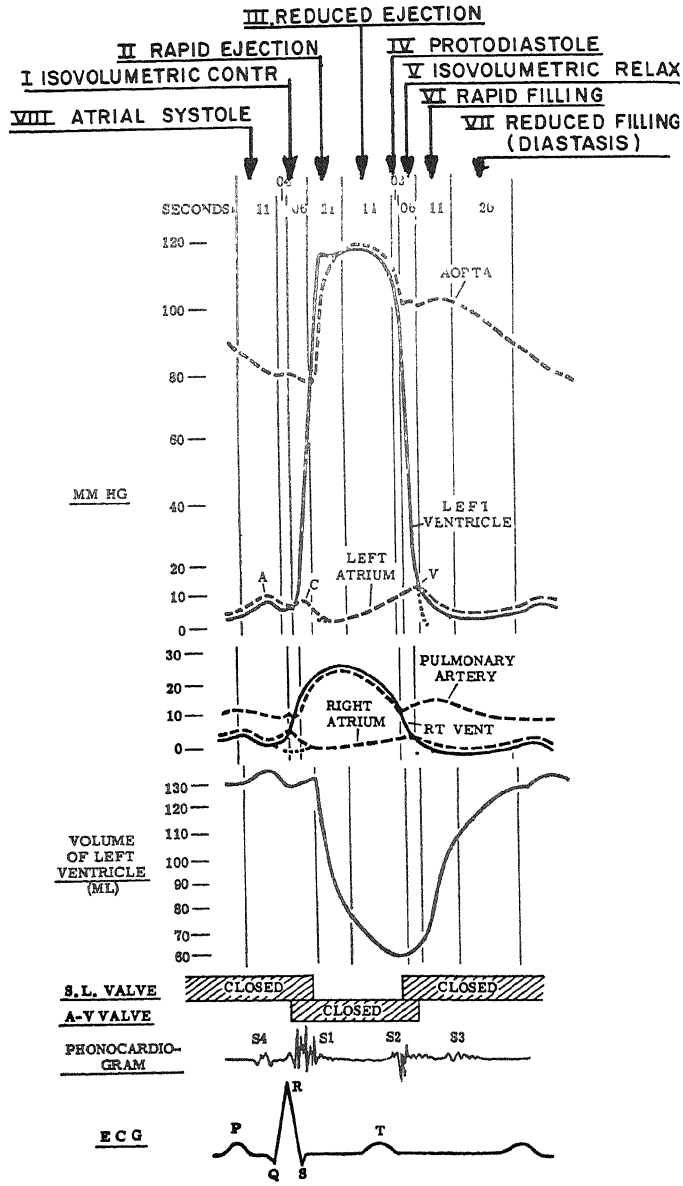


Fig 3 110 Events of the cardiac cycle *Upper* left heart *lower* right heart (Redrawn from Selkurt and Bullard 1966)

systole, aortic pressure is exceeded by left ventricular pressure at the time of greatest acceleration of aortic flow

III Reduced Ejection During rapid ejection, blood flow into the aorta (and pulmonary artery) has exceeded runoff into the peripheral arteries and pressure continues to rise As runoff reaches equilibrium with flow into the great arteries, the pressure curve reaches a maximum This is the beginning of the reduced ejection period Runoff now exceeds cardiac output and the pressures begin to fall The pressure in the left ventricle falls slightly below that in the root of the aorta, although forward flow continues as indicated by progressive decrease in ventricular volume

IV Protodiastole This marks the beginning of ventricular relaxation and is a point on the

ventricular pressure curves which is often difficult to identify The pressures in the ventricles continue to fall below those in the aorta and pulmonary artery A brief retrograde flow occurs, closing the semilunar valves This marks the end of protodiastole and the beginning of the next phase

V Isovolumetric Relaxation With closure of the semilunar valves, the incisura in the arterial pressure curves is seen, marking the beginning of this phase Once again, the ventricles are closed chambers and myocardial relaxation continues, causing a steep fall in intraventricular pressures but little alteration in ventricular volume When the intraventricular pressures fall below those in the atria, the atrioventricular valves open, marking the end of this period During this phase, the ventricles have just completed ejection of perhaps 50 to 65% of their diastolic volumes, and thus remain partially filled with blood

VI Rapid Filling Beginning with the opening of the atrioventricular valves, ventricular volumes increase rapidly as blood which has accumulated in the atria under increasing pressure flows quickly into the relaxed ventricles During this phase the ventricular volume curve ascends steeply at first and then begins to level off The end of this phase is not clearcut as it merges with the next phase At the transition between this and the following phase the usually inaudible third heart sound may be recorded in the phonocardiogram

VII Reduced Filling (Diastasis) This is a period of slower filling during which blood continues to flow into both atria and ventricles as into a common chamber. It is terminated by onset of atrial systole

VIII Atrial Systole This period is marked by the onset of the atrial contraction ("A") wave which is reflected in the ventricle since the two chambers are now continuous This atrial systolic pressure wave peaks and begins to decline, finally being interrupted by the onset of ventricular isovolumetric contraction, completing the cardiac cycle

During the preceding ventricular contraction, the atrial pressure undergoes a temporary abrupt decline as a result of an artefact imposed by cardiac movement However, as blood pours into the cavity of the left atrium from the veins, the atrial pressure rises continuously to the end of the isometric relaxation period forming the "V" wave This atrial blood which has accumulated during the previous ventricular contraction is now transferred to the ventricle Ventricular filling begins at the time the intraventricular pressure drops below the intra-atrial pressure and the AV valves reopen

The period of rapid ventricular filling, i e , the early diastolic inflow period is marked by a continuous decline in ventricular pressure and in the "V" wave of the atrial pressure curve (the so-called "Y" descent) Inflow into a ventricular cavity is regulated by mechanisms operating in both its systole and diastole Since no blood enters the ventricles until it has passed through the atria, it is important to know their means of filling It is not possible to measure flow into the atria, but since the right atrium is continuous with the venae cavae without intervening valves, vena caval flow should indicate right atrial flow Caval flow into the right atrium shows three phases It is reduced nearly to zero during atrial contraction During ventricular diastole inflow is considerable, but the volume of atrial inflow is greater during ventricular systole when the AV valves are closed and no blood can enter the ventricle, than during ventricular diastole when AV valves are open and atria and ventricles form a common cavity Atrial inflow during ventricular contraction is particularly effective in the presence of tachycardia in which the time available for diastolic inflow is shortened The piston-like downward movement of the atrioventricular junction during ventricular systole enlarges the atria and vena cava and attracts blood to them. The filling of the atria during ventricular systole depends on the quantity and pressure of blood available in the venous reservoir and the vigor with which ventricular systole moves the AV junction. Thus, ventricular myocardial contraction is a very important means of actively filling the atria during systole and, therefore, in determining the extent of ventricular filling during the rapid filling phase

Finally, attention is directed to the dotted and abruptly dipping line which represents an alternate decline of ventricular pressure, and at times indicates a negative intraventricular pressure has developed (*vis a fronte*, p 3-200) Such a dip in the pressure curve of ventricular relaxation, although recorded for many years and considered an artifact, is now given more credence since *1)* it can be recorded with a micromanometer (which is supposedly immune to

artifacts), 2) recent experimental evidence indicates that the relaxing ventricle can develop suction under certain artificial experimental conditions For example, in the beating mammalian heart, a moderate volume of fluid can be drawn into the ventricular cavity in the absence of a positive ventricular filling pressure and in the presence of an empty ventricle or one with some systolic residual volume The time available during which suction could act is approximately 0 1 sec , i e , the earlier part of the rapid inflow period The quantitative contribution of the ventricle to ventricular filling under normal conditions is still unknown and, hence, the physiological significance of diastolic suction for the return of blood to the heart cannot be evaluated Wiggers, in summarizing the problem, states, "crucial evidence is still required that the small remnant of elastic recoil still operative at the end of relaxation can create sufficient suction to be of significance in filling the normally beating heart Dynamically it must be shown that the concordant declines of atrial and ventricular pressures are due to a more rapid rate of ventricular relaxation than of filling from the atrium " However, there are significant restoring forces during the early part of diastole which probably aid diastolic filling (Tyberg et al , 1970)

Atrial Contraction. The dynamic importance of atrial contraction on ventricular filling has been much debated The older view that the chief function of the atria is filling the ventricles is not now acceptable since it is known that such filling occurs chiefly during the earlier portion of diastole when the difference between atrial and ventricular pressure is maximal However, in hearts slowed by vagal stimulation to separate atrial and ventricular contraction and in hearts with 2 1 or more block, it can be demonstrated that atrial contraction has a small but significant effect on the ventricular volume curve obtained with a cardiometer In open-chest dogs with surgically induced complete atrio-ventricular block and a fixed, slow ventricular rate, there is an increase in left ventricular end-diastolic pressure, stroke volume, stroke work and aortic systolic pressure when atrial systole precedes ventricular systole by an interval of 0 085 to 0 125 seconds, indicating enhanced ventricular filling through atrial contraction (see also p 3-25) Finally, atrial contraction and relaxation are instrumental in bringing about closure of the atrioventricular valves (see below)

Pressure Pulse Values. Aortic pressure throughout the cardiac cycle is uniformly higher than in the pulmonary artery The peak aortic pressure during systole of 120 mm Hg is approximately five times that of 25 mm Hg in the pulmonary artery, the relative end-diastolic pressures are 80 mm Hg versus 10 mm Hg, the relative pulse pressures are 40 mm Hg vs 15 mm Hg The diastolic pressure in both cavities is quite low with a small pressure gradient from the left ventricle (0 to 5 mm Hg mean diastolic, 5 to 12 mm Hg end-diastolic) to the right ventricle (0 to 3 mm Hg mean diastolic, 0 to 5 mm Hg end-diastolic) The atrial pressure values are only a few mm Hg and are uniformly somewhat higher in the left atrial (mean 2 to 12 mm Hg) than in the right atrial (mean 0 to 5 mm Hg) curve

Asynchronisms of Pressure Pulses While dynamic events on the two sides of the heart are generally similar, there is considerable asynchronism and some difference in duration of parts of the cardiac cycle The onset of contraction of the right atrium precedes that of the left atrium while the onset of contraction of the right ventricle follows that of the left ventricle Right ventricular ejection begins earlier and is completed later than left ventricular ejection (Fig 3 111)

Electrocardiogram vs. Onset of Pressure Events In man the interval between the onset of electrical and mechanical activity in the left atrium (rise of P wave vs onset of rise of atrial A wave) is of the order of 0 08 sec , for the right atrium, the interval approximates 0 06 sec For the left ventricle, the interval between the onset of ventricular depolarization (Q wave)

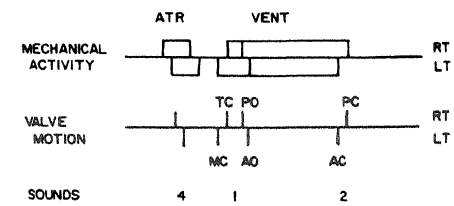


Fig 3 111 Sequence of valve motion as related to mechanical activity during the cardiac cycle Right heart events above the line left heart events below the line ATR atrium VENT, ventricle RT, right LT left MC mitral closure TC tricuspid closure PO pulmonic opening AO aortic opening AC, aortic closure PC pulmonic closure 4 1 2, fourth, first second heart sounds Duration of isometric contraction for each ventricle is indicated in each ventricular rectangle by a verticle line (Redrawn from Lewis, 1962)

and onset of left ventricular contraction (onset of pressure rise) approximates 0 04 sec , for the right ventricle, the interval is about 0 05 sec The T wave has a variable relation to the end of systole but terminates usually before the incisura of the aortic pressure curve

Aortic Flow. The relationship between left ventricular pressure pulse and aortic flow in a dog is seen in Figure 3 112 The flow rate at the root of the aorta increases rapidly during the phase of rapid ejection when most of the systolic output of the heart is expelled, then flow falls off during the phase of reduced ejection and reverses at the time of protodiastole During diastole the aortic flow remains at about zero However, forward flow continues in the more peripheral arteries under the impetus of its initial momentum and the elastic recoil of the larger arteries

The Ventricular Pressure-Volume Diagram (Work Diagram). For a complete representation of mechanical activity of the heart, the coordinates of pressure, volume, and the time must be employed In Figure 3 110, pressure and volume curves are plotted separately against time The relationships between ventricular pressure and volume changes during the cardiac cycle can be displayed as functions

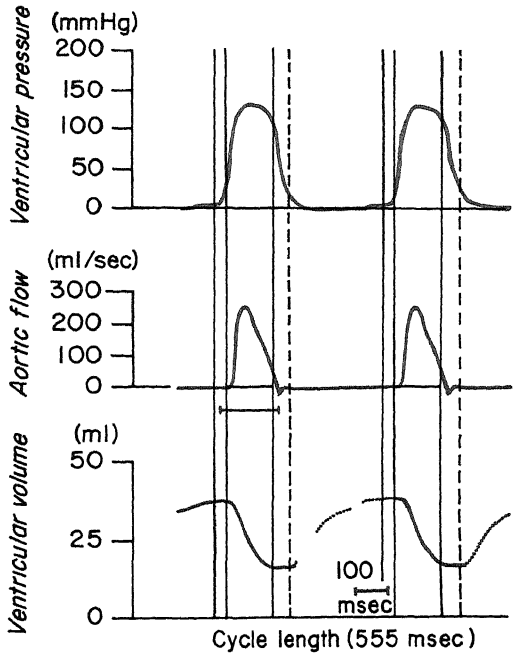


Fig 3 112 Aortic flow in relationship to ventricular pressure and volume changes during the cardiac cycle (23-kg dog) (Redrawn from Jacob 1968)

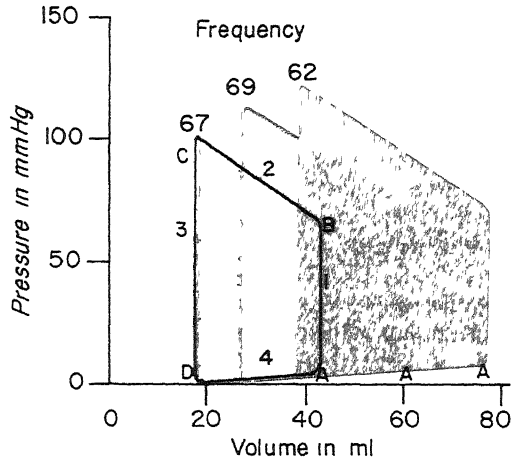


Fig 3 113 Ventricular pressure volume diagrams Data from a 20-kg dog The letters and numbers identify the following phases of the cardiac cycle *A* onset of ventricular contraction (beginning of isovolumetric contraction) level of end diastolic pressure and volume *B* onset of ventricular ejection (end of isovolumetric contraction) *C* end of ventricular ejection (onset of isovolumetric relaxation) level of end-systolic pressure and volume *D* end of isovolumetric relaxation (beginning of diastolic period) *1* isovolumetric contraction period *2* ventricular ejection (for purposes of simplification the pressure during ejection is depicted as continuously increasing) *3* isovolumetric relaxation period *4* diastole The effect of increasing end-diastolic pressure and volume (*A*) is illustrated Note the increase in area of the work diagram as the levels of end-diastolic pressure and volume (*A*) increase (Redrawn from Bauereisen 1970)

of one another In Figure 3 113, pressure and volume values during the cardiac cycle are plotted at three different rates of cardiac filling The area bounded by the sides of each quadrangle represents work (pressure × volume) The effect of increasing end-diastolic pressure and volume on stroke work is indicated (Frank-Starling mechanism)

The Pressure-Volume-Time Diagram. The relations between pressure and volume changes with time during a ventricular cycle are diagrammed three-dimensionally in Figure 3 114 The temporal changes in pressure and volume are diagrammed on the left and base planes, respectively, and the work diagram on the right plane The three dimensional curve in the center represents power (rate of work or work per unit time), since the time course of the work diagram can be determined

MOVEMENTS OF HEART VALVES

Atrioventricular Valves (Tricuspid and Mitral). The valve leaflets or cusps, three in

number on the right and two on the left side, are attached by their bases to the fibrous rings surrounding the atrioventricular openings Their free margins are connected through delicate tendons (chordae tendineae) to the papillary muscles which prevent inversion of the valves into the atrium during ventricular systole The chordae tendineae are tightened at the commencement of systole by the contraction of the papillary muscles The leaflets are composed mainly of a double layer of the endothelial lining of the heart, strengthened by fibers of connective tissue Their attached bases are thicker and contain more connective tissue, small blood vessels and delicate strands of smooth muscle The latter, however, play no part in valve closure which is effected, as mentioned above, in a passive manner

The Mechanism of Valve Closure According to traditional concept, an advancing gush of blood thrusts the AV valves widely open during early ventricular diastole During atrial systole the leaflets do not lie back against the ventricular wall but occupy a midposition as a result of two opposing currents. The inflowing blood pressing upon their atrial surfaces keeps them open, while eddies reflected in the reverse direction from the ventricular walls strike their ventricular surfaces and tend to close them Thus, they float in a position of delicate balance When, as a result of the fall in intra-atrial

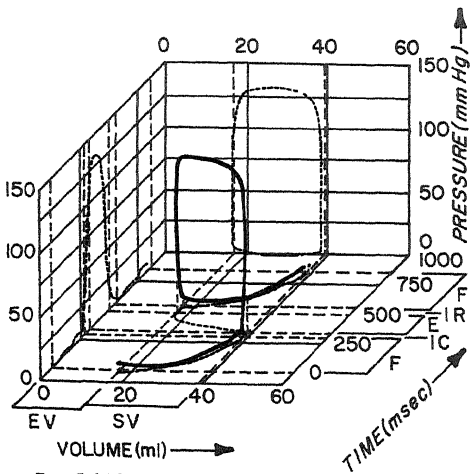


Fig 3 114 Pressure-volume-time diagram of the left ventricle Base volume changes with time Left vertical plane pressure changes with time Posterior vertical plane pressure volume diagram (Redrawn from Jacob 1968)

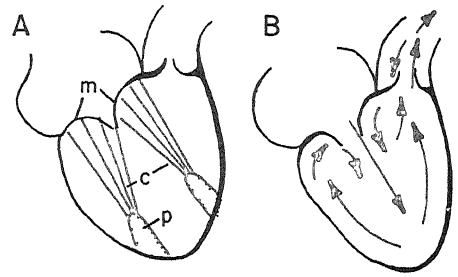


Fig 3 115 Two diagrams showing mechanisms concerned in valve closure *A* showing relations of papillary muscles and chordae tendineae to valve flaps *B* partial closure due to eddy formation (After Wiggers 1954)

pressure at the end of atrial systole, the incoming jet is diminished in force and finally ceases, the back eddies persisting for a brief space and being unopposed, approximate the valves or bring them gently into apposition (Fig 3 115) The importance of atrial relaxation in closing the valves has been shown in open-chest dogs with surgically induced atrio-ventricular block, where a negative atrioventricular pressure difference was demonstrated during atrial diastole The fact that the first heart sound has been found to be significantly delayed in the absence of atrial systole supports this view that normally the atrioventricular valves are closed or nearly so, before ventricular systole They are not, however, firmly closed This is effected by the rise in pressure in the ventricle when it contracts Dean (1916) has shown by attaching a hair to the septal leaf of the valve and recording its movements, that if ventricular systole does not follow almost instantly upon the cessation of the flow of blood from the atrium, the valves start to reopen In instances, therefore, in which ventricular systole is delayed, that is, when the A-V interval is prolonged, the reopening of the valves proceeds for an appreciable time Then, when ventricular contraction occurs, a small amount of blood regurgitates into the atrium before the valves are swung closed by the rising intraventricular pressure The backward flow of blood through the orifice may then give rise to a murmur just preceding the first heart sound (presystolic murmur)

These concepts have been derived mainly from observations and direct photography of valve action in isolated hearts and from correlation of heart sound with mechanical and electrical events of the cardiac cycle However, cinefluorographic observations of the motions of

the mitral valve cusps by means of attached radiopaque markers in intact anesthetized dogs, showed the edges of the valve cusps to be only slightly separated during ventricular filling Since in both instances, it is not possible to rule out certain experimental artifacts, no choice can be made at present between the concept of wide valve excursion in isolated hearts and restricted valve movements in intact animals

The Semilunar Valves. The dynamics of aortic and pulmonary closure are essentially the same in principle as those described for the AV valves The valves and sinuses of Valsalva form three small pockets open toward the arterial lumen Back eddies which are set up during the ejection phase of systole prevent the contact of the valves with the arterial wall When ejection ceases, the centripetal currents carry the valves into apposition and firm closure is effected by the higher pressure at this time upon their arterial surfaces

HEART SOUNDS

The sounds produced by the heart during a cardiac cycle consist of waves of greatly varying frequency Because of the intervening tissue, the intensity of the sounds is generally decreased and some components are accentuated and others depressed

Graphic registration of the heart sounds has been used to give a permanent record for study and to permit precise relations to be established between them and the events of the cardiac cycle, as recorded by other graphic methods such as the electrocardiogram and intracardiac and intravascular pressure curves

For obvious reasons, sounds picked up from the surface of the thorax are not a true representation of the actual vibrations of the cardiac valves and walls Thus, attempts have been made to move the pick-up device closer to the source of sound by placing microphones in the esophagus, by suturing microphones to the myocardium in experimental animals, by applying suction microphones directly to the surface of the human heart, and finally, by introducing the microphone into the cardiac chambers themselves The latter technique aids in locating the site of production of normal heart sounds and serves as a valuable diagnostic tool in disease states in locating the origin of abnormal sounds and murmurs However, use of data obtained thereby must be tempered by knowledge of the fact the catheter mounted device may itself, through the motion imparted to it in the cardiac cavities, produce sounds which are recorded as naturally occurring

Characteristics of Heart Sounds. Two distinct sounds are heard by direct auscultation and occasionally a third is also audible during a cardiac cycle A fourth sound is revealed by graphic methods The first two sounds occur mainly during ventricular systole, the other two sounds take place during ventricular diastole (Fig 3 110)

Origin of Heart Sounds and Time Relations These vibrations arising in the heart and great vessels are fundamentally the result of sudden displacement of blood (acceleration) or abrupt stoppage of blood (deceleration), and turbulence

The temporal relations of human heart sounds to central hemodynamic events are illustrated schematically in Figure 3 110 The large vibrations of the first sound begin with the rapid rise in intraventricular tension and by the beginning of ejection they have already reached a maximum The beginning of the second sound coincides with the trough of the incisura of the aortic pressure The third sound occurs near the end of the rapid inflow phase of diastole of the ventricles The fourth sound begins near the middle of the atrial pressure wave and extends to the isometric phase of ventricular contraction

First Heart Sound The first sound is of relatively long duration, soft in quality and low in pitch The second sound is shorter, sharper and of higher pitch These characteristics can be imitated vocally by the syllables "lub" and "dup" separated by a brief pause The two heart sounds mark the beginning and end of ventricular systole and the determination of the interval between their commencements (as determined by the phonocardiogram) is a reliable method for arriving at the length of ventricular systole in man although intracardiac phonocardiographic techniques indicate a time lag of the order of 0 02 seconds between the closure of the atrio-ventricular valves and the beginning of the first heart sound (see below) The pause between the end of the second sound and the beginning of the first coincides with ventricular diastole

The *First Sound* is heard most clearly and at maximum intensity over the fifth left inter-

costal space, i e , with an area centered over the "apex beat " Here the mitral element of the sound predominates Any abnormal sound produced at the tricuspid valve is detected by listening over the lower end of the sternum

Four components of the first heart sound have been identified The first component is associated with the earliest phase of ventricular contraction which shifts blood toward the atrioventricular valves, approximating them It precedes elevation of intraventricular pressure The second component begins with closure and tension of the atrioventricular valves as intraventricular pressure begins to rise The third component begins as intraventricular pressure starts to exceed arterial pressure with the onset of ejection It is associated with opening of the semilunar valves and abrupt distention of the arterial trunks The fourth component probably results from flow turbulence in the arterial trunks during early ejection

The record of the first sound is composed chiefly of a series of from 9 to 13 vibrations, and has a duration of from 0 9 to 0 16 sec The vibrations are in general irregular, which places the sounds in the category of noises rather than of musical tones The frequency is low varying from 33 to 110/sec

The Second Sound results from the vibrations set up in the blood column and arterial walls as the aortic and pulmonary valves are placed under tension following their closure The duration of the second sound is about 0 10 sec It commences about 0 09 sec after the summit of the T wave of the electrocardiogram The average frequency is about 50 per sec In the pulmonary area (along the left upper border of the sternum) two components of the second sound can normally be heard ("split second sound") Phonocardiographic records show that the first of the two components is synchronous with the dicrotic notch of the carotid arterial pressure pulse and therefore represents aortic valve closure, the second component is caused by closure of the pulmonary valve The relative delay of pulmonary vs aortic valve closure is enhanced during inspiration when the increased venous return to the right heart prolongs right ventricular ejection time, the situation is reversed during subsequent expiration, when the blood ejected by the right ventricle reaches the left side of the heart and causes prolongation of left ventricular ejection, moving aortic valve closure closer to pulmonary valve closure Thus the splitting of the second heart sound widens during inspiration (average 0 05 sec) and narrows or even disappears during expiration (average 0 02 sec) The relationships of valvular opening and closure, mechanical activity of the atria and ventricles and heart sounds are shown in Figure 3 111

The Third Heart Sound Sometimes a faint third sound is heart in normal hearts which follows the second sound by about 0 08 sec and lasts for about 0 04 sec It is heard at the apex and is commonly found in young adults It may be made to appear or is intensified by procedures which increase the venous flow into the atria, e g , exercise, recumbent position, etc Several explanations have since been given to account for it By intracardiac phonocardiography it was demonstrated that the sound coincides with rapid, early ventricular filling and is recordable from the ventricular cavity and the ventricular surface but not from the atrium It is suggested that the sound is produced by the vibrations of the ventricular wall as blood rushes into the ventricle

The Fourth Heart Sound, as a rule, is inaudible in normal subjects Often, however, the phonocardiogram records a few small pre-systolic vibrations, usually consisting of two components The first of these coincides with the atrial "a" wave and, by intracardiac phonocardiographic techniques, it can be recorded from within the atria, indicating that it is caused by atrial contraction The second component follows the peak of atrial contraction and can be recorded more often from the ventricles than from the atria, suggesting that it is due to the impact of blood, expelled from the atria, against the ventricular wall There is also evidence, however, from studies on horses (Patterson et al, 1965) that this late component results from transient reversal of the pressure gradient between atria and ventricles at the end of atrial systole, resulting in presystolic tensing and perhaps closure of the atrioventricular valves

Variations in the Intensity of the Heart Sounds. It is the general belief that the first sound varies in intensity with the force of ventricular systole and the loudness of the second sound with the height of the arterial blood pressure The intensity of the first sound is directly related to the rate of the pressure rise within the ventricle during the isometric period The intensity of this sound is not dependent upon

the volume of the systolic discharge but rather upon the diastolic pressures in the pulmonary and systemic circuits. Wiggers found that when the heart was slowed and the systolic discharge consequently increased but the diastolic pressure lowered, the intensity of the first heart sound was reduced while acceleration of the heart (reduced systolic discharge with raised diastolic pressure) increased its intensity. These experiments together with the observations that the first heart sound is believed to be largely on a valvular basis and is greatly influenced by the length of the P-R interval, strongly suggest that the intensity of the sound is ultimately conditioned by the position of the AV valves at the onset of ventricular systole.

The intensity of the second sound, aortic or pulmonary, is increased by an elevation in the systemic or pulmonary pressures, respectively. Among the cardiovascular conditions associated with intensification of the second sound are, mitral stenosis and failure of the left ventricle which raise the pulmonary arterial pressure, and arterial hypertension which raises the aortic pressure.

5 / REGULATION OF THE HEART

Regulation of cardiac output is effected through mechanisms which operate within the myocardium itself (*intrinsic* or *autoregulation*) and factors outside the heart (*extrinsic regulation*) These mechanisms adjust cardiac frequency and stroke volume, thus altering cardiac output The functional properties of myocardial fibers affected include force, velocity, duration, and extent of contraction Conductivity is adjusted to maintain proper timing of atrial and ventricular systole and the temporal sequence of ventricular activation Finally, the output of the left and right hearts is coordinated so that they remain equal

Intrinsic Regulation of Ventricular Function

Control mechanisms within the myocardium itself make it possible for isolated myocardial preparations to adapt to changes in work load In dogs with denervated hearts, cardiac output adjusts appropriately during exercise and rest The mechanisms responsible reside within the heart and thus are called *intrinsic* The best known such adaptive response is that to changes in resting myocardial fiber length, the Frank-Starling mechanism Since this involves a change in myocardial fiber length, it has been termed *heterometric autoregulation* The term *homeometric autoregulation* is used to designate other intrinsic adaptive mechanisms which do not involve myocardial fiber length

HETEROMETRIC AUTOREGULATION

As described for isolated papillary muscle (p 3-110 et seq) and the entire isolated heart (p 3-119 et seq), increasing fiber length or diastolic volume of the heart, within physiological limits, increases contractile force and ventricular stroke work Thus, the denervated heart can respond appropriately to alterations in filling pressure, duration of diastole (filling time), and its own distensibility

HOMEOMETRIC AUTOREGULATION

Myocardial tension, frequency of contraction, and temperature influence contractile strength for any given fiber length

Myocardial Tension. The isolated heart may maintain its increase in stroke volume in response to an increase in venous return even after its diastolic volume has returned to the control level Similarly, if the ejection resistance (afterload) is increased in such a preparation, the stroke volume at first decreases, but can return to control values Then the same stroke volume is maintained against an increased aortic pressure This response, which is not dependent on changes in fiber length, is sometimes called the Anrep phenomenon Anrep (1912) first demonstrated that when aortic resistance was elevated in the heart-lung preparation (see p 3-121, Fig 3 120), ventricular volume first increased, then returned to its control value, presumably maintaining comparable stroke volume against elevated pressure Increased myocardial wall tension triggers some unknown mechanism which can elevate contractile strength Possibly, the initial stretch alters the viscous properties of the series and parallel elastic elements of the myocardium, allowing alterations in the degree of contractile filament overlap (Braunwald et al , 1968) Local release of norepinephrine by the increased stretch has also been suggested as a possible factor, and in conscious dogs infusion of isoproterenol (a β-adrenergic receptor stimulant) increases the positive inotropic response of the left ventricle (Bugge-Asperheim and Kiil, 1969)

Rate and Rhythm Effects. The strength of myocardial contraction is markedly influenced by the time interval between beats The earliest demonstration of this effect was that of Bowditch (1871), who described the staircase (Treppe) phenomenon in the frog ventricle, after a period of rest, subsequent contractions increase stepwise in strength until a plateau is reached

In mammalian ventricular muscle a premature contraction is weaker than the preceding beats Then, for several regular beats following the premature beat, the contractions are increased in strength This potentiation cumulates if premature contractions are continually repeated This phenomenon has been called *postextrasystolic potentiation* or *premature activation potentiation* (*prematurity potentiation*) The latter terms are preferable In ventricular muscle of most mammals, rate of tension development, peak tension developed, and maximum shortening increase with frequency over the physiological range This is termed *frequency potentiation* Frequency potentiation requires many beats to reach a steady state while prematurity potentiation is almost complete with the first potentiated contraction

Details on the influence of the time interval between beats on papillary muscle contraction are illustrated in Figure 3 116 According to the analysis of Blinks and Koch-Weser (1961), two interval-dependent processes influence myocardial contractility in opposite directions

Negative inotropic effect of activation (*NIEA*) is a temporary effect tending to decrease the strength of contraction in subsequent beats Each beat has a relatively large NIEA which disappears rapidly within a few seconds after contraction ceases

Positive inotropic effect of activation (*PIEA*) is a temporary effect tending to increase the strength of contraction in subsequent beats Each beat has a smaller PIEA than NIEA and the effect disappears slowly with a half time of decay to the order of 0 5 to 1 0 min

Both NIEA and PIEA are cumulative The strength of a given contraction is that of the rested-state contraction plus cumulated PIEA, minus cumulated NIEA Under steady-state conditions at regular frequency the PIEA and NIEA produced by each beat decay during each interval between beats Each beat produces more NIEA than PIEA, but the NIEA decays much more rapidly Premature beats occur in the presence of more of both PIEA and NIEA than regular beats, but NIEA predominates, causing premature beats to be weaker than regular beats The longer lasting PIEA is predominant at the end of the longer interval between regular beats

In Figure 3 116, various effects of interval changes on tension development of a mammalian papillary muscle are shown In record A, the points are steady-state tensions developed at various frequencies (interval between contractions of 0 2 to 300 sec) This illustrates frequency potentiation In B, prematurity potentiation is illustrated The contractions following the premature contractions are stronger for several beats owing to PIEA, which finally decays and the contraction returns to control levels In C, shortening of the interval from 20 to 0 63 sec produces initially potentiation due to prematurity At first, this is followed by a progressive decline in contractile force for a few beats, caused by NIEA Then *Treppe* (staircase phenomenon) follows owing to accumulation of the longer lasting PIEA When the frequency of contraction is again slowed, the time course of PIEA decay can be determined (C)

The effect of PIEA accumulation can be produced by so-called *paired pacing*, in which the heart is driven by paired pulses (Kelly, 1969, Cranefield, 1965) The time interval between paired stimuli is short so that the second stim-

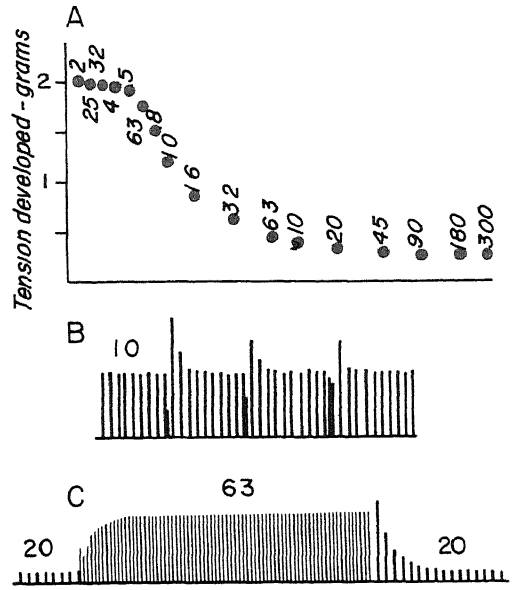


Fig 3 116 Tension responses in an isolated papillarly muscle as the interval between contractions is varied *A* graph of the steady-state tensions developed at the between contraction time intervals (seconds) indicated above each point *B* the influence of premature contractions in a preparation stimulated at 1 0-sec intervals *C* the effects of changing the interval between contractions from 20 sec to 0 63 sec (Redrawn from Koch-Weser and Blinks 1963)

ulus of each pair occurs just after the absolute refractory period following the first induced contraction The second stimulus produces negligible mechanical effect, but the following beat is potentiated (PIEA) When the premature activations are continually repeated, PIEA cumulates and cardiac action improves even in the presence of congestive heart failure The use of paired pacing in cardiac patients with failure, while effective, has not produced impressive clinical results

The mechanisms accounting for these inotropic effects are not known There is increasing evidence that PIEA of frequency potentiation is caused by increasing Ca^{++} influx and greater availability of Ca^{++} at contractile sites The NIEA seen in the premature beat may be caused by changes associated with the relaxation process and membrane repolarization

The inotropic effects of activation described in the foregoing are characteristic of the ventricle of most mammalian species The interval-strength relations are very different in most types of atrial muscle and in ventricular muscle of various species

Temperature. In isolated preparations of mammalian heart, hypothermia augments contractile strength In contrast, moderate hypothermia (body temperature 25° to 30° C) and pyrexia (body temperature 41° to 43° C) decreased ventricular contractile capacity in intact dogs (Goodyer, 1965) The direct stimulant effect of hypothermia observed in isolated preparations is replaced by a depressant effect in intact hearts A possible explanation for this is that cooling depresses cardiac sympathetic tone in the intact heart, since adrenergic blocking agents largely prevent this effect Pyrexia, on the other hand, enhances ventricular contractile capacity even after autonomic blockade The thermal limits of the human and dog heart are between 26 ° and 44° C, near these limits myocardial contractility and conduction are depressed

Extrinsic Regulation

NERVOUS CONTROL OF THE HEART

The Heart Rate. The average rate in adult man is around 70/min , but there is a rather wide variation between individuals, a rate considerably below or above this average being not uncommon Muscular training tends to reduce the cardiac rate, athletes not infrequently having a pulse rate between 50 and 60 On the other hand, a rate between 80 and 90 is sometimes seen in other healthy persons The rate diminishes progressively from birth, when it is around 130/min , to adolescence, but increases slightly again in old age Among physiological conditions which temporarily increase the heart rate are *muscular exercise, emotional excitement, and high environmental temperature* It also increases somewhat during *digestion* The rate is lowered during *sleep* (55 to 60) Among pathological conditions which cause an increase in cardiac rate are *hemorrhage, surgical shock, hyperthyroidism, fever* (an increase of 10 beats/1° F rise in temperature), and certain *cardiac arrhythmias,* e g , paroxysmal tachycardia, atrial fibrillation, etc

Tachycardia and *bradycardia* are general terms used to denote, respectively, any considerable increase in heart rate above, or reduction below the normal average

The heart, as we know, beats rhythmically after its complete separation from the central nervous system, but in the intact animal this automatic action is under the continuous influence of nervous impulses Where these were blocked experimentally with appropriate drugs in man, the heart beat intrinsically at a mean rate of 107 5/min (standard deviation 9 3) at rest (Jose and Taylor, 1969) The nervous mechanism comprises groups of nerve cells in the medulla—the *cardiac centers,* various *afferent pathways* along which impulses are conveyed to these centers from numerous regions of the body, and the *vagus* and *accelerator* or *augmentor* nerves which transmit impulses from the centers to the heart

Innervation of the Heart. The vagus nerves are cardio-inhibitory This action was discovered by the Weber brothers in 1845 They convey fibers, belonging to the parasympathetic division of the involuntary nervous system, from a center in the medulla (*cardio-inhibitor center*) to the special tissues of the heart The medullary center was located by Miller and Bowman (1915) in the dorsal nucleus of the vagus situated in the floor of the fourth ventricle. Weak electrical stimulation of this area produced slowing of the beat, and stronger currents, complete arrest of the heart The cardiac fibers of the vagus separate from the trunk of the nerve in the neck between the origins of its superior and inferior laryngeal branches Inter-

mingling with fibers of the accelerator nerves, they enter into the formation of the deep and superficial cardiac plexuses, whence they are continued to the atrial muscle Here they make connection with ganglion cells Postganglionic fibers pass to the specialized tissue of the sinoatrial and atrioventricular (AV) nodes where they are prolonged between the muscle fibers They form a rich plexus and are seen to end as ring-shaped or club-shaped structures (boutons") upon the fibers of the specialized tissue Many postganglionic fibers do not enter the nodal tissues but terminate in the atrial myocardium proper Some vagal fibers are distributed to the ventricular myocardium as well as to coronary arteries With vagal stimulation, negative inotropic effects on both atria and ventricles can be demonstrated In dogs, greater cardiac acceleration follows section of the right vagus than of the left, and stimulation of the right vagus has a greater tendency to cause sinus arrest than does left vagal stimulation Left vagal stimulation causes atrioventricular conduction block preferentially, however, with stimulation, left and right vagal effects on AV conduction time were about equal (Hamlin and Smith, 1968)

Accelerator or Augmentor Nerves The accelerator fibers were described by von Bezold in 1863 They belong to the thoracolumbar division of the involuntary nervous system and arise from cells situated in the lateral horns of the upper thoracic segments of the spinal cord These cells constitute a spinal cardio-accelerator center The preganglionic fibers enter the gangliated cord of the sympathetic to connect with cells in the *inferior*, *middle*, and *superior cervical ganglia* In many animals and also often in the human subject the inferior cervical and the first thoracic ganglia are fused into an irregularly shaped structure called the *stellate ganglion* from which accelerator fibers pass directly to the heart The heart also receives accelerator fibers *directly* from the sympathetic chain as far down as the fourth or fifth thoracic ganglion In order, therefore, to remove all accelerator influence from the heart it is necessary to interrupt these connections as well as to remove the stellate ganglia The axons of the cells of the cervical ganglia (postganglionic fibers) form the *inferior*, *middle*, and *superior cardiac nerves* (Fig 3 117) These fibers, especially those forming the nerves of the right side, terminate in the sinoatrial node Those of the left side are distributed mainly to AV node and bundle According to Nonidez (1943), sympathetic efferent fibers which reach the heart are contained mainly in the middle cardiac nerve The superior cardiac nerve is distributed to the large arteries at the base of the heart while the inferior cardiac nerve is mainly afferent The spinal accelerator center is subordinate to higher centers

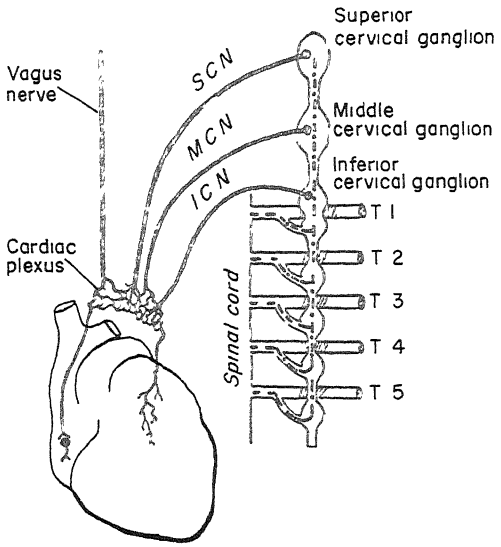


Fig 3 117 Diagram of the cardiac nerves Broken lines = preganglionic sympathetic fibers

In dogs, stimulation of the left stellate ganglion produces a greater effect on ventricular contractility (augmentor fibers) than on heart rate (accelerator fibers) while with stimulation of the right stellate ganglion these relative effects are reversed (Levy et al , 1966) Selective right and left sympathetic denervation studies also indicate that the SA node and atria are innervated primarily from right, and the ventricles from left sympathetic fibers (Priola, 1969)

Tonic Action of Nerves The functional properties of the heart affected by extrinsic nerve impulses, chemical mediators, hormones, drugs, etc , are contractile strength, rate, conductivity, and excitability Classically, the following qualitative terms have been applied to these effects *chronotropic action* (*chronotropy*), influence on heart rate, *inotropic action* (*inotropy*), influence on contractile strength, rate of

pressure development (dP/dt), ejection velocity, *dromotropic action* (*dromotropy*), influence on conduction velocity, and *bathmotropic action* (*bathmotropy*), influence on excitability or irritability (stimulation threshold)

An increase in action is termed *positive* and a decrease *negative* Thus, a *positive inotropic* effect is one which increases strength of contraction, a *negative chronotropic* effect is one which slows heart rate

The sympathetic fibers stimulate the heart and thus affect the aforementioned properties in a positive sense, the parasympathetic fibers inhibit the heart and affect these properties in a negative sense The chemical mediators of these autonomic effects, norepinephrine (sympathetic) and acetylcholine (parasympathetic), act on the specific adrenergic (sympathetic) and cholinergic (parasympathetic) receptor sites in effector cells Of the two types of adrenergic receptor sites recognized, α and β, the heart contains chiefly β-receptors Activation of the α-receptors causes vasoconstriction but few primary cardiac effects There is evidence for the presence of vasoconstrictor α-receptors in the coronary arteries (Feigl, 1968) and negative chronotropic α-receptors in the SA node (Priola, 1969, James et al, 1968) Activation of the β-receptors causes vasodilatation and positive chronotropic and inotropic cardiac effects (see also p 3–115)

The Tone of the Vagus At rest, the vagus nerves exert a continuous restraint upon the action of the heart Thus, the vagus, or rather the cardioinhibitory center, possesses tone, impulses passing from it in a continuous stream to the heart This effect can be readily demonstrated in animals by cutting or freezing the nerves The heart's action then immediately becomes greatly accelerated The increase in rate following the removal of the vagal influence also occurs though the stellate ganglia have been previously excised, the result therefore cannot be due to an increased action of the accelerator nerves The tonic action of the vagus nerves may be annulled by atropine, the heart rate then increasing to 150 or 180/min The difference between this rate and the normal resting rate of 70/min , therefore, represents the vagal effect which is being constantly exerted under ordinary circumstances Various conditions, physiological and pathological, alter the tone of the vagus center The tone is naturally higher in some species, e g , the horse, which is capable of feats of endurance, than in others, e g , the domestic rabbit It also shows individual variations in man, athletes usually showing a higher tone than those who lead sedentary lives

Vagal tone is apparently reflex in nature and dependent upon afferent impulses flowing to the vagus center especially along the sinus and aortic nerves Section of these nerves causes an increase in heart rate and little further acceleration occurs, as a rule when the vagi themselves are subsequently severed

Sympathetic Tone The existence of a tonic action of the sympathetic nerves has been postulated on the basis of experiments showing that excision of the stellate ganglia was followed by a slowing of the heart rate A criticism of these experiments is that they were not performed under truly basal conditions—i e , in conscious, fasting animals resting comfortably, with external visual and auditory stimuli reduced to a minimum and, in addition, apprehension of the animal eliminated through previous training to lie quietly during observation of the heart rate Yet only under these circumstances can the presence or absence of true accelerator tone (as opposed to accelerator nerve activity under various experimental conditions) be ascertained Murphy (1942) studied dogs in this manner The average "resting" heart rate before training was 90 to 112/min After several weeks of training, the resting rate slowed and stabilized at 47 to 57/min Cardiac sympathectomy was then performed Following recovery of the animals from this procedure the basal rate was 50 to 58—i e , not different from that observed prior to interruption of the accelerator fibers It was concluded that in the truly resting state the accelerator nerves are not in tonic activity, but that the accelerator mechanism is more or less constantly involved in adapting the heart rate to the changing bodily conditions

Cardiac Reflexes. A variety of reflexes affect both cardiac activity and vascular tone simultaneously They are therefore discussed together (see chap 8)

Under ordinary conditions, the activities of the cardio-inhibitory and cardio-accelerator centers which result in the continuous discharge of impulses along the corresponding cardiac nerves are in turn dependent to a very large extent, if not entirely, upon the reception of impulses by afferent paths In other words,

the maintenance of the tone of the centers, and so of the normal resting rate of the heart, and the alterations in rate which occur under various physiological conditions are in large measure either reflex in nature or due to impulses received from cerebral centers The impulses which stream into the nervous centers arise in all parts of the body, the heart itself included By these influences the tone of either center may be exalted or depressed, and corresponding changes produced in the cardiac rate If the cardiac vagus on one side be cut and its central portion (i e , the end leading to the brain) stimulated, a reflex through the cardioinhibitory center and the opposite vagus occurs which alters the cardiac rate The nature of the change in rate—whether acceleration or inhibition—which will result from stimulation of the central end of the vagus or of most other afferent nerves cannot always be foretold Reflex slowing of the pulse can usually be demonstrated in the human subject by pressure upon the eyeball at the outer canthus (oculo-cardiac reflex), carotid sinus, or by the stimulation of nasal branches of the fifth nerve Stimulation of afferent fibers in the respiratory passages as by the inhalation of irritating vapors, e g , anesthetics, is particularly likely to cause reflex inhibition of the heart Extrasystoles and bradycardia have been demonstrated electrocardiographically in man during abdominal operations, the irregularities being the consequence, apparently, of visceral stimulation Excitation of the central end of various peripheral nerves, e g , the sciatic, causes reflex changes in the pulse rate In these last instances acceleration is more readily obtained than inhibition The irradiation of impulses on to the cardiac centers from the cerebral centers, e g , from the motor area at the commencement of muscular exercise or from regions concerned with emotional manifestations, are held responsible for the changes in pulse rate which occurs under these conditions

The pulse rate is generally inversely related to the arterial blood pressure, a rise or fall in pressure causing, respectively, a decrease or increase in heart rate These adjustments are believed to be subserved by *1*) a reflex whose afferent limb is constituted of afferent vagal fibers (aortic nerve) ending in the aortic arch and heart, 2) a reflex in which the sinus nerve forms the afferent limb These mechanisms have been considered in chapter 8 Presumably such changes in heart rate are brought about not simply by an increase or a decrease in tone of one or the other cardiac centers, but by reciprocal variations in the tone of both For example, the slowing of the heart which results from a rise in arterial pressure is much less pronounced if impulses from the cardio-inhibitory center have been prevented from reaching the heart by section of the vagi After removal of the stellate ganglia on the other hand, the cardiac response to a rise in blood pressure is reduced to a lesser extent Other mechanisms are concerned, since a fall in blood pressure causes an increase in the rate of the heart even after it has been completely denervated

CHEMICAL CONTROL OF THE HEART

The *adrenal medullary catecholamines* exert positive inotropic chronotropic, dromotropic, and bathmotropic actions on the myocardium and induce glycogenolysis These are β-adrenergic receptor effects, it has been suggested that the β-receptor may be adenyl cyclase (p 3-114), since a close correlation has been found between the positive inotropic and chronotropic effects of catecholamines and their adenyl cyclase activating potency Adenyl cyclase catalyzes the conversion of adenosine triphosphate (ATP) to 3',5' cyclic adenosine monophosphate (3',5' AMP), this 3',5' cyclic AMP then promotes the conversion of inactive phosphorylase *b* to active phosphorylase *a* Enhanced glycogenolysis and increased energy supply in cardiac muscle is the result Epinephrine stimulates 3',5' cyclic AMP production and Ca^{++} mobilization in the heart Both are necessary for the increase in phosphorylase *a* activity and the resultant enhancement of glycogen breakdown it catalyzes Increased Ca^{++} permeability at the cell surface occurs and 3',5' cyclic AMP increases efficiency of Ca^{++} mobilization from stores in the sarcoplasmic reticulum Thus, the positive inotropic effect of catecholamines can be attributed to an increase in Ca^{++} availability Such a mechanism of activation may operate in other cells and it has been proposed that 3',5' cyclic AMP and Ca^{++} are two interrelated intracellular messengers which function in both muscle and secretory cells (Rasmussen, 1970) (see also p 3-34) The way in which this system could operate in heart muscle is illustrated in Figure 3 118

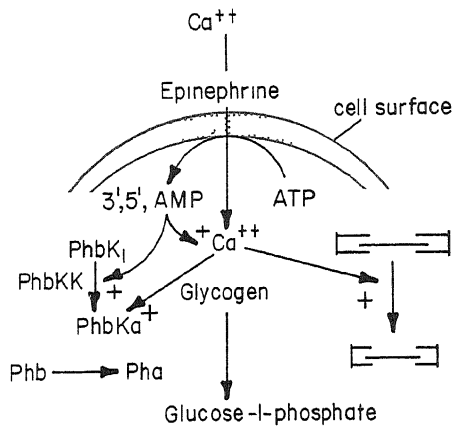


Fig 3 118 Hypothetical mechanism of epinephrine inotropic and glycogenolytic effects in heart muscle Epinephrine stimulates 3' 5' cyclic AMP production and calcium mobilization These sequentially activate phosphorylase *b* kinase kinase (PhbKK) and inactive phosphorylase *b* kinase (PhbKi) to active phosphorylase *b* kinase (PhbKa) Inactive phosphorylase *b* (Phb) is then converted to active phosphorylase *a* (Pha) and glycogenolysis is increased The positive inotropic effect results from greater availability of Ca^{++} made possible by increased Ca^{++} uptake by the cell This is induced by epinephrine and by more efficient Ca^{++} mobilization from the sarcoplasmic reticulum stimulated by cyclic AMP (Modified from Rasmussen 1970)

Glucagon causes a similar activation of glycogenolysis and a positive inotropic effect, possibly by a similar mechanism It does not act at a β receptor site, but appears to stimulate cardiac adenyl cyclase at a different receptor site This suggests there are at least two receptor sites affecting catalytic activity of the cyclase, one responding to adrenergic β stimulation and one to glucagon (Mayer, 1970)

Adrenal cortical hormones have positive inotropic effects as do *angiotensin* and *serotonin* *Thyroxin* possesses positive chronotropic actions, but whether or not it is directly inotropic is uncertain

Hypoxemia, *hypercapnia*, and *acidosis* of sufficient degree depress myocardial contractility

Cardiac Output and Cardiac Work

The most important external manifestation of the heart is its output of blood and the work it performs These measurements are obtained by the Fick procedure, by the dilution principle, and in some instances, by the foreign gas methods (pp 3-158 to 3-161)

DEFINITIONS AND GENERAL CONSIDERATIONS

The output of the heart per beat is spoken of as the *systolic discharge* or the *stroke volume*, and the output per minute as the *minute volume* The value of the latter is simply the product of the stroke volume and the pulse rate (pulse rate $\times$ stroke volume = minute volume), the minute volume divided by the pulse rate, therefore, gives the stroke volume The quantity of blood ejected by each beat of the left ventricle in the average healthy man during rest is from 70 to 80 ml An equal quantity is, of course, discharged at the same time by the right ventricle, making a total for the whole heart of from 140 to 160 ml The contents of the left ventricle are ejected against a much higher mean arterial pressure than the contents of the right, the mean pressure in the pulmonary artery being about one-sixth of that in the aorta The minute volume is expressed in terms of one ventricle The output of one ventricle obviously represents the quantity of blood flowing through the lungs, or through the systemic vessels, during the same period An adult of sedentary occupation pumps at least 5500 liters of blood through his body daily—from left to right ventricle through the systemic vessels, and from right to left through the lungs

The portion of the energy output of the left ventricle that appears as useful work can be roughly calculated from the product of mean aortic blood pressure and cardiac output The work done by the right ventricle can be similarly calculated, except that the mean pressure is that in the pulmonary artery The end diastolic pressure in the left or right ventricle should be subtracted from the measured aortic pressure or pulmonary arterial pressure since the end diastolic ventricular pressure does not have its origin from ventricular systole Preferably, the transmural pressure should be subtracted, but this is difficult to obtain in man Ordinarily the end diastolic pressure is small, approximating up to 5 mm Hg in the right ventricle and up to 10 mm Hg in the left Yet under some conditions of heavy ventricular loading such as an arteriovenous fistula, or congestive heart failure, the value may reach 30 to 40 mm Hg, and hence, the correction is important The simpler calculation also assumes that the velocity factor is but a very small part of the external work of the heart

Normally, this is of the order of 1 to 2% When cardiac output is quite large in the presence of a decreased systemic pressure, or especially if aortic or pulmonary stenosis occurs, the kinetic factor can be up to 50% of the cardiac work (see also pp 3-96 to 98)

To compare better the cardiac output and cardiac work in different individuals, the values are generally expressed in terms of surface area and are spoken of as flow or work indices Thus, the cardiac index (CI), stroke volume index (SVI), cardiac work index (CWI), and stroke work index (SWI) are obtained by dividing the cardiac output, stroke volume, cardiac work, and stroke work by the surface area of the body An average surface area is 1 7 m^2 Normal values for the right and left heart obtained in man during catheterization studies are in Table 3 13 The reason for relating cardiac output to body surface area in man is that, just as in the case of basal metabolic rate, the correlation with body size is better when based on body surface area than when based on body weight Actually, the basic determinants of cardiac output are the oxygen requirement of the tissues, the volume of oxygen carried per unit volume of blood, and the percentage extraction of oxygen from the blood in tissue capillaries (i e , the arteriovenous oxygen difference) The oxygen requirement depends on metabolic rate and blood oxygen content on the concentration of respiratory pigment (e g , hemoglobin) The arteriovenous oxygen difference (i e , percentage oxygen extraction by tissues) is governed by the blood-tissue oxygen gradient, the capillary surface area available for gaseous exchange, and the time the blood remains at these surfaces In mammalian species (mouse to elephant) "metabolic weight" is an exponential function of body weight (e g , $kg^{0.75}$) In larger mammals (dog to horse) cardiac output is linearly related to body weight, the output in liters per minute approximates 0 1 times the body weight in kilograms [more precisely cardiac output (liters/min) = (0 1017) × (body weight in $kg^{0.99}$)(Patterson et al , 1965b) However, if other species are considered (e g , fish and crustaceans), body weight and body surface area are no longer useful as a reference for comparison of oxygen requirements as related to circulatory flow rates of blood Much higher relative blood flow rates are required in those species which have blood with low oxygen carrying capacity (Dejours et

Table 3 13
Hemodynamic studies in man of the right and left heart during rest and exercise both at sea level and at high altitude

Condition	Treadmill Exercise kg m / min /m^2	Heart Rate	Mean BP[a] mm Hg	CI[a] liters/ min	TPR[a] (Press)/ (Flow)	CWI,[a] kg m / min /m^2	SVI[a] ml	SWI[a] gm m / min /m^2	O_2 usage liters/ min /m^2	AV O_2 ml /100 ml
Right heart (average surface area, 1 83 m^2)										
Rest		73	16	3 4	2 6	0 63	47	8	0 132	4 1
Exercise	Leg movements	152	35	9 2	2	3 94	67	26	1 047	11 32
Left heart (average surface area, 1 74 m^2)										
Lima										
Rest		81	92	2 8	19	3 6	33	41		
Exercise	704-986	163	109	8 1	8	12 1	51	75	1 353	16 9
Left heart (average surface area, 1 62 m^2)										
Morococha										
Rest		83	101	4 2	15	6 0	51	72	0 209	5 7
Exercise	490-580	156	118	7 5	10	12 3	49	79	1 072	13 9
Exercise	695-749	175	125	15 2	5	26 4	88	151		

[a] BP, blood pressure, CI, cardiac index, TPR, total peripheral resistance, CWI, cardiac work index, SVI, stroke volume index, SWI, stroke work index

al , 1970) Thus, for comparative purposes, as between species, the relationship between circulatory requirements as reflected by cardiac output, must include consideration of arterial oxygen concentration Similarly, in man, if the oxygen carrying capacity of the blood is diminished (as with the low hemoglobin concentrations seen in anemia), cardiac output must increase to maintain optimal tissue oxygen supply for metabolic needs In man the ratio of cardiac output (Q) to oxygen consumption Vo_2 approximates (in absolute terms) 20 1 (e g , $\dot{Q}/Vo_2$ = 6000 ml /300 ml) If the oxygen carrying capacity is reduced, as in anemia, this ratio increases as blood flow increases to carry the required oxygen The general trend of change is similar to that observed in fish and crustacea, and further displays the importance of oxygen requirement as a basic determinant of circulatory volume flow

OUTPUT OF THE HEART NORMAL STANDARDS, PHYSIOLOGICAL VARIATIONS

The minute output of the heart under basal conditions[1] for the average size male adult is about 5 5 liters, and the CI is 3 2 liters The CI is slightly lower for the female and considerably higher for children Cardiac output can be quite variable in the same person from day to day unless extreme precautions are used to maintain the resting state

Conditions in Which Cardiac Output Is Not Changed. The cardiac output is apparently the same when a person is asleep as when he is awake and in the horizontal position It is unaffected by menstruation, by metabolic acidosis produced by the infusion of ammonium chloride or lactic acid, and by moderate variations in the external temperature, although it may be increased up to 30% at environmental temperatures above 30° C , or by low temperatures that are associated with shivering The native indigenous to high altitude undergoes considerable acclimatization by changes in his cardiovascular and respiratory systems, so that at altitudes of, at least, 15,000 ft , the cardiac output is normal as are his heart rate and systemic blood pressure and systemic arteriovenous O_2 However, the slightest exertion will increase his cardiac output which will be considerably higher in the sitting rather than in the reclining position, whereas in persons indigenous to sea level, such change in posture at sea level may decrease cardiac output

Conditions in Which Cardiac Output Decreases.

Normal The cardiac output is decreased about 30% in changing from the recumbent to the upright posture

Pathological States Atrial fibrillation and a very rapid heart rate may reduce cardiac output up to 50% In complete heart block and with an efficient myocardium, the output may be approximately normal, but when this exists in the presence of coronary sclerosis, myocardial degeneration, or after open heart surgery (cutting of the atrioventricular bundle), the output may be greatly reduced In other subjects, although the output at rest is normal, the inability of the heart to accelerate properly prevents an adequate output during muscular exertion In congestive heart failure, cardiac output very often is reduced, and the increase in cardiac output which normally follows muscular effort does not occur or is slight This type of failure is to some extent compensated for by a greater increase in the systemic arteriovenous O_2 difference than occurs in healthy persons doing a similar amount of external work

Conditions Which Increase Cardiac Output. Cardiac output may be increased 50 to 100% by anxiety and excitement and, at least, 30% by eating (during first 3 hr) Anemia, pregnancy (in later months), metabolic alkalosis produced by infusion of sodium bicarbonate, and low oxygen or high CO_2 in the inspired air are conditions which increase cardiac output up to 100% In maximum exercise the cardiac output may be increased five, seven, or even ninefold

Effects of Drugs upon the Cardiac Output. *Adrenaline* and *histamine* increase the oxygen consumption and the cardiac output, the effect upon the minute volume is, however, proportionately greater than that upon the oxygen consumption *Acetylcholine*, whose effects in general are very evanescent, causes a slight increase in the minute volume A decided rise in the cardiac output is produced by *nitrites* which, like acetylcholine, cause arteriolar dilation The increase in the cardiac output caused by nitrites is probably a compensatory effect of the lower peripheral resistance, whereby the

[1] Body recumbent, at rest and at 20° C for 15 to 20 min and 12 hr after food and drink

blood pressure is maintained near the normal level *Digitalis* produces no immediate effect upon the minute volume in normal persons In subjects with congestive heart failure and atrial fibrillation the output is as a rule, though not invariably, increased as the cardiac condition improves *Atropine* which increases the heart rate to 150 or 180 beats/min does not increase, as a rule, the cardiac output *Alcohol* in moderate dosage (35 ml) causes no more than a slight rise in the cardiac output

REGULATION OF CARDIAC OUTPUT AND CARDIAC WORK DURING EXERCISE AND EXCITEMENT

In strenuous exercise, the cardiac output may reach 35 to 45 liters/min (Table 3 13) or increase 7 to 9 times, while the body oxygen usage may increase 10 to 12 times

Heart Rate. At the transition from rest to work, the pulse frequency rises very rapidly with heavy exercise, reaching levels of 160 to 180/min During short exhaustive work, heart rates as high as 240 to 270 have been recorded in normal young people This increase in heart rate is attributed to some form of autonomic control, but the mechanisms have not been established The suspect Bainbridge reflex probably does not apply since this requires for its operation cardiac distension from increased venous return As we shall see later, the heart is more apt to decrease in size The tachycardia during exercise is attributed to a reflex through the carotid sinus and aortic baroreceptors resulting from a reduced arterial pressure as the result of vasodilation in the active muscle However, the systemic blood pressure rises, not falls, during exercise Finally, tachycardia is explained as related to direct regulation from the higher centers

Ventricular Pressure-Volume Relations. It has long been believed that the output of the heart is governed by its venous blood supply, and that the increase in minute output during muscular work requires a regulating mechanism by which venous return is augmented To effect this, the venous blood supply and pressure in the central veins must increase to such a height that the increased venous flow and ventricular filling pressure will dilate the heart and increase the length of its fibers during each diastole, and by this stretch cause a greater ventricular response in the following systole

A very large amount of experimental work has stressed this viewpoint Howell and Donaldson (1883) first showed the intrinsic adaptation of the heart of the dog to an artificially increased venous return Frank (1895) in the frog heart, came to the conclusion that the magnitude of ventricular contraction is determined by the intracavity volume at the end of systole In 1914, Patterson, Piper, and Starling measured the pressure and volume of both ventricles with a cardiometer in a heart-lung preparation of the dog, while the venous return was passively and progressively increased by elevating a blood infusion reservoir connected to the right atrium As the atrial pressure or ventricular filling pressure rose, the heart dilated and with each systole ejected a greater stroke volume at a greater ventricular pressure This relationship, called "Starling's Law of the Heart" or the "Frank-Starling" mechanism, indicated in a more formalized way that "the energy of contraction, however measured, is a function of the length of the muscle fibers," before contraction (Fig 3 119*A*) (pp 3–88 to 92, and 3–44) As these investigators indicated, this relationship is little more than the application to cardiac muscle of what was already known about skeletal muscle by Blix in 1891 and Fick in 1882 This relationship of ventricular pressure and volume has been confirmed for the whole heart in the open-chest dog and for the left ventricle in the open- and closed-chest dog (Fig 3 119*B*).

Because of the difficulty of estimating volume of a single ventricle and the relative ease of pressure measurement, attempts have been made to correlate mean atrial pressure or end-diastolic pressure of the associated ventricle with the stroke volume or stroke work A useful experimental preparation for this type of study by Sarnoff et al (1958) is in Figure 3 120 Sarnoff has found that the curves expressing the relation of mean atrial pressure and stroke work are shifted in one direction by conditions improving myocardial performance (stimulation of cardiac fibers from the stellate ganglia, reflex stimulation of the myocardium by a decrease in carotid sinus perfusion pressure, injection of epinephrine), while unfavorable conditions such as vagal stimulation, increase in carotid sinus perfusion pressure and coronary insufficiency shift the curves in the opposite direction These ventricular function curves provide important objective evidence

that ventricular control can involve changes in myocardial contractility, a concept now widely accepted (Fig 3 119*C*, *D*)

As one approaches the normal state, however, the evidence is equivocal for this relationship The peripheral venous pressure in man during heavy work (2 8 liters oxygen/min) can increase up to 7 mm Hg Variations of filling pressure, however, cannot be deduced from peripheral venous pressure because of the unknown pressure gradient from the periphery to central veins In the dog exercising on the treadmill, in the human following injection of a small amount of epinephrine or exercising mildly while lying down, there may be no direct correlation between the increased cardiac output or cardiac work and the atrial or ventricular end diastolic pressure Furthermore, intracardiac end-diastolic pressure measurements alone do not necessarily represent the effective transmural pressure, since it was shown in the closed-chest dog that under certain experimental conditions a rise in ventricular end-diastolic may be accompanied by a

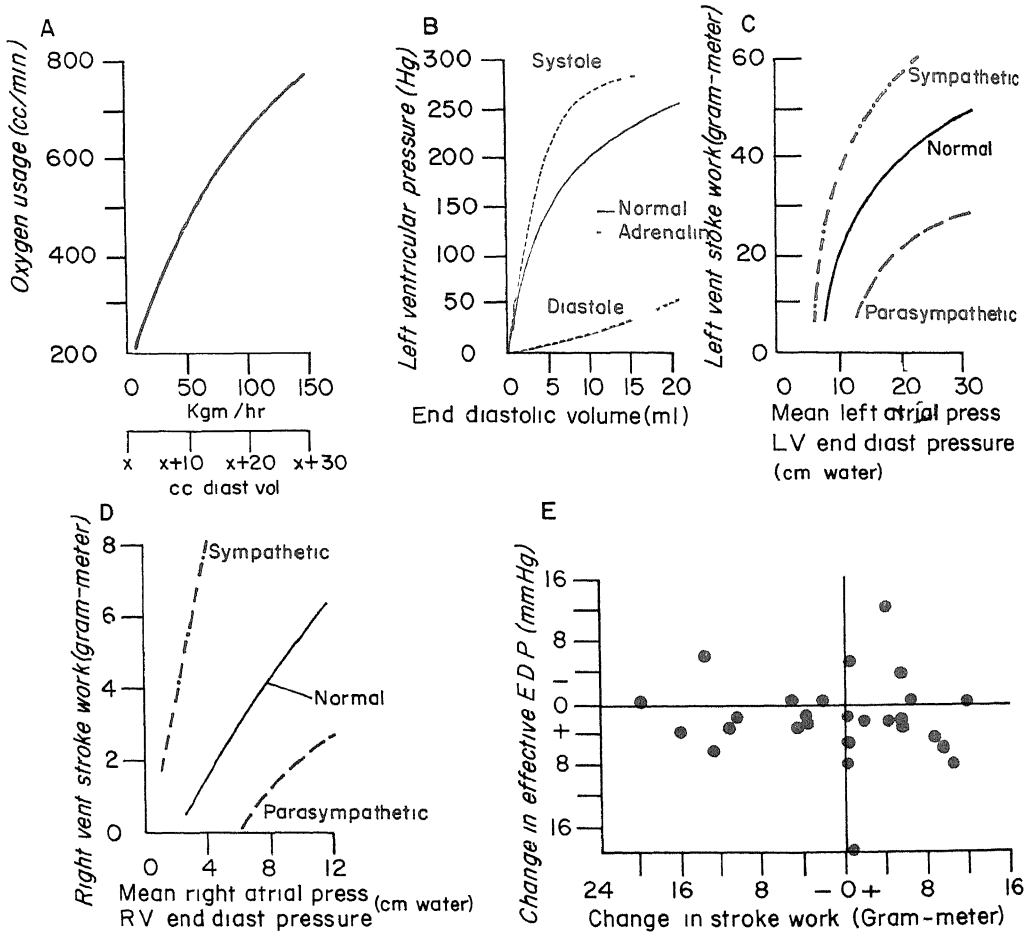


Fig 3 119 Performance characteristics of the heart *A* Relation of O_2 usage to cardiac work and diastolic volume i e the energy output of the heart utilizing the heart lung preparation ' (Starling and Visscher 1926) *B* pressure-volume function curves of the left ventricle in the open-chest dog *C* Effect of sympathetic and parasympathetic cardiac nerve stimulation on the relation between mean left atrial pressure and left ventricular stroke work and relation between left ventricular end-diastolic pressure and left ventricular stroke work (Sarnoff 1960) *D* response of right ventricular stroke work to mean right atrial pressure and to ventricular end-diastolic pressure and the effects of autonomic stimulation The parasympathetic response is presented only for direction and does not indicate magnitude (Sarnoff 1960) *E* change in effective end-diastolic pressure (*EDP*) vs stroke work in the normal dog before and within 1 min after exercise (From Gregg et al 1955)

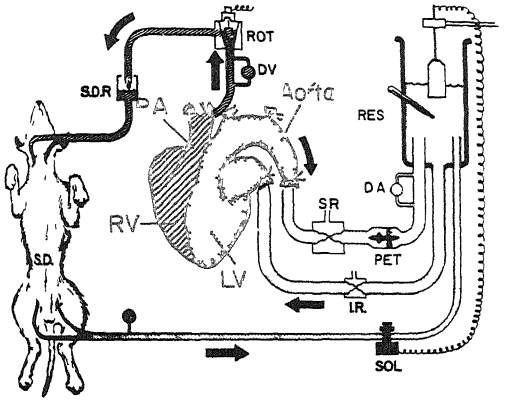


Fig 3 120 Modern version by Sarnoff of isolated heart or heart-lung preparation The coronary venous return is passed through a donor dog to maintain the blood at a normal biochemical environment before it is infused into the coronary system of the isolated heart In older versions of the Starling isolated heart or heart-lung preparation in which the coronary venous blood circulating through the myocardial wall was simply reoxygenated and returned to the coronary system progressive deterioration in performance characteristics developed within 60 to 90 min *SD* support dog *SDR* support dog reservoir *SOL* solonoid valve electrically operated by microswitch at top of reservoir float *SR* air-filled Starling resistance *PET* Plotter electroturbinometer *DA* arterial densitometer *RES* reservoir *IR* water-filled inflow Starling resistance *DV* venous densitometer *ROT* rotameter (Sarnoff et al 1958)

concomitant increase in the intrapericardial pressure On the other hand, in subjects with an artificial pneumotherax, a linear relation was found between effective or transmural atrial pressure and stroke volume with an increase of 2 mm Hg providing an increase in stroke volume of 40% In humans whose autonomic nervous system has been inhibited by a ganglionic-blocking drug elevation of the left ventricular end-diastolic pressure, induced by blood transfusion, was accompanied by an increase of the cardiac output and stroke volume This relationship also apparently holds in patients with high output failure and very large ventricular volumes

Ventricular Volume and Stroke Volume. As already indicated, comparisons of direct measurements of ventricular volume with pressure or with stroke volume or stroke work have not been made in the normal state in which the aggregate complications of venous resistance, intrathoracic pressure, vascular volume, changing heart rate, neurogenic factors, and humoral influences exist However, highly suggestive evidence has been obtained by estimating the ventricular volume in systole and diastole The essential findings follow,

In man and the dog, the left ventricular diastolic volume and stroke volume increase in the recumbent position as compared to the more or less upright position Concurrently, the central blood pool or thoracic blood volume increases by about 400 ml The cardiac output and heart rate change in opposite directions on alteration of position, i e , in the supine position the cardiac output increases and the heart rate decreases Correspondingly, assuming the upright position is accompanied by a reduction in heart size and stroke volume

In the resting dog, an increase in venous return induced by a rapid infusion of blood or by displacing blood from the splanchnic area into the thorax by abdominal compression increases the left ventricular dimension without significantly altering stroke volume or stroke work Excitement or startle or the anticipation of exercise just before treadmill activity can lead to a transitory increase in stroke volume with enlargement of the left ventricle

Nonexhaustive exercise by healthy dogs and normal untrained supine human subjects may be accompanied by increased cardiac output, primarily resulting from tachycardia with little or no progressive increase in stroke volume as the level of exertion increases Stroke volume during such exertion may only occasionally exceed the resting recumbent values in healthy dogs or untrained human subjects, although with exercise performed in the upright position (when the resting stroke volume is smaller), a definite increase in stroke volume accompanies the increasing heart rate and cardiac output Concurrently, the left ventricular dimension of the dog's heart remains the same or becomes smaller However, trained athletes and patients with chronic volume loads, such as in valvular insufficiency, may increase cardiac output with a significant augmentation of stroke volume and a lesser degree of tachycardia The relatively small hearts of untrained persons may show a tendency to reduce their diastolic size during exercise but the end-diastolic size of the large hearts of athletes gets much smaller The larger capacity of athletes, as compared to untrained subjects, for increasing cardiac output during exercise is explained by the larger stroke volume

Evidence for increased contractility of the

myocardium during exercise is in the increasing rate of tension development, rate of blood ejection, and rate of relaxation as seen in the ventricular pressure curves and the aortic flow curves

The preceding evidence suggests, but does not offer quite certain proof, that the fundamental property of cardiac muscle of a relation between end ventricular diastolic size and energy release is not dominant, but rather that in the normal dog and man, an increase in cardiac output during exercise is achieved mainly by an increase in heart rate with a constant or slightly increased stroke volume A little reflection will indicate, however, that this is a self-regulating mechanism and cannot alone explain the observed changes in cardiac output When the heart accelerates, the output per beat can be kept constant only if the venous inflow is adequate, only under such circumstances can the minute volume be increased

The contribution which cardiac acceleration can make toward the minute volume is, therefore, strictly limited, for rates from 180 to 200 are about the maximums to which the healthy heart can be speeded up This is only about 2½ to 3 times the normal rate With a constant stroke volume, therefore, cardiac acceleration could not increase the minute volume more than about 3 times Yet reliable data have shown that the cardiac output of untrained man doing strenuous exercise may increase ninefold (from 5 to 6 liters/min at rest to 35 to 45 liters) and the oxygen consumption 12 times (from 250 ml oxygen at rest to 3000 ml) In this circumstance, it must be concluded that the stroke volume increases 2 to 3 times (even at a maximum pulse rate) to account for up to 50% of the cardiac output increase This is borne out by some of the data in Table 5 13

In summary, it can be stated that both changes in stroke volume and heart rate contribute significantly to regulation of cardiac output in man, but no firm conclusion can be drawn as to the relative role that stroke volume, heart rate, and diastolic and systolic volumes play under various conditions of stress

Body Arteriovenous (AV) O_2 Difference. Since the body oxygen usage can increase 12 times in exercise, it is important to know the mechanisms by which this amount of oxygen is supplied to the tissues The AV O_2 difference is generally calculated from the oxygen usage divided by the cardiac output, the former being determined by collection and oxygen analysis of the expired air and the cardiac output by the dilution technique, which does not require cardiac catheterization The oxygen extraction rises but considerably more slowly than does the cardiac output, reaching in very heavy exercise a calculated maximum of 13 to 16 ml /100 ml of blood Since most of the increase in blood flow is presumably through the exercising muscle, this means that the oxygen extraction at times must be very nearly complete in the latter, a situation similar to that which has been observed in the heart muscle with maximal activity In well-trained subjects, it appears that the rise in AV O_2 difference, as well as heart rate, are somewhat smaller, while the stroke volume is larger (see p 3-221)

Other Hemodynamic Variables in Exercise In addition to the heart rate and stroke volume alterations occurring in exercise, many other changes are involved which may be mild or almost insignificant in light exercise, or which may be massive is strenuous activity (Table 3 13) In the presence of heavy exercise and a very large increase in cardiac output, the CWI of the left ventricle may increase four times, the systemic blood pressure may increase considerably, the arterial pressure pulse may increase from 50 to 100 mm Hg, the total peripheral resistance may decrease greatly Presumably, the latter results mainly from vasodilation in the metabolically active beds for, since splanchnic blood flow in man is decreased in the presence of an elevated systemic blood pressure, resistance in the splanchnic bed must be increased In the pulmonary circulation, the arterial pressure rises moderately, while the resistance decreases or undergoes little change Finally, as pointed out earlier, the central venous, atrial, and ventricular end-diastolic pressure do not necessarily increase The alterations in many hemodynamic parameters during exercise is indicated in Figure 3 121

The Role of Catecholamines. The importance of the sympathetic nervous system in the regulation of the heart's performance has already been mentioned on several occasions throughout this section The effects of sympathetic nerve stimulation are in all likelihood mediated through the release of norepinephrine from the nerve endings Evidence for this view may be summarized as follows

The mammalian heart, including the human, like several other tissues and organs possessing

a rich adrenergic innervation, contains significant stores of norepinephrine The norepinephrine content of the tissues is closely related to the number of adrenergic fibers supplying them It disappears completely following postganglionic sympathetic denervation Similarly, the administration of reserpine results in practically complete disappearance of norepinephrine from the heart, and, in animals so treated, the positive chronotropic effect of accelerator nerve stimulation is blocked The cardiac effects of stellate ganglionic stimulation can be mimicked by norepinephrine infusion Finally, during stellate ganglionic stimulation, norepinephrine can be recovered from coronary venous blood, in an amount proportional to the intensity of the stimulus

REGIONAL DISTRIBUTION OF OXYGEN CONSUMPTION AND BLOOD FLOW

The preceding discussion indicates that despite much study of long familiar functions, their operation in the adjustments of oxygen and cardiac output to exercise is poorly understood The response of oxygen usage and blood flow, of course, represents the average requirements of organs and regions of the body The advent of implantable sonar and electromagnetic flowmeters now makes it possible to measure blood flow simultaneously in a limited

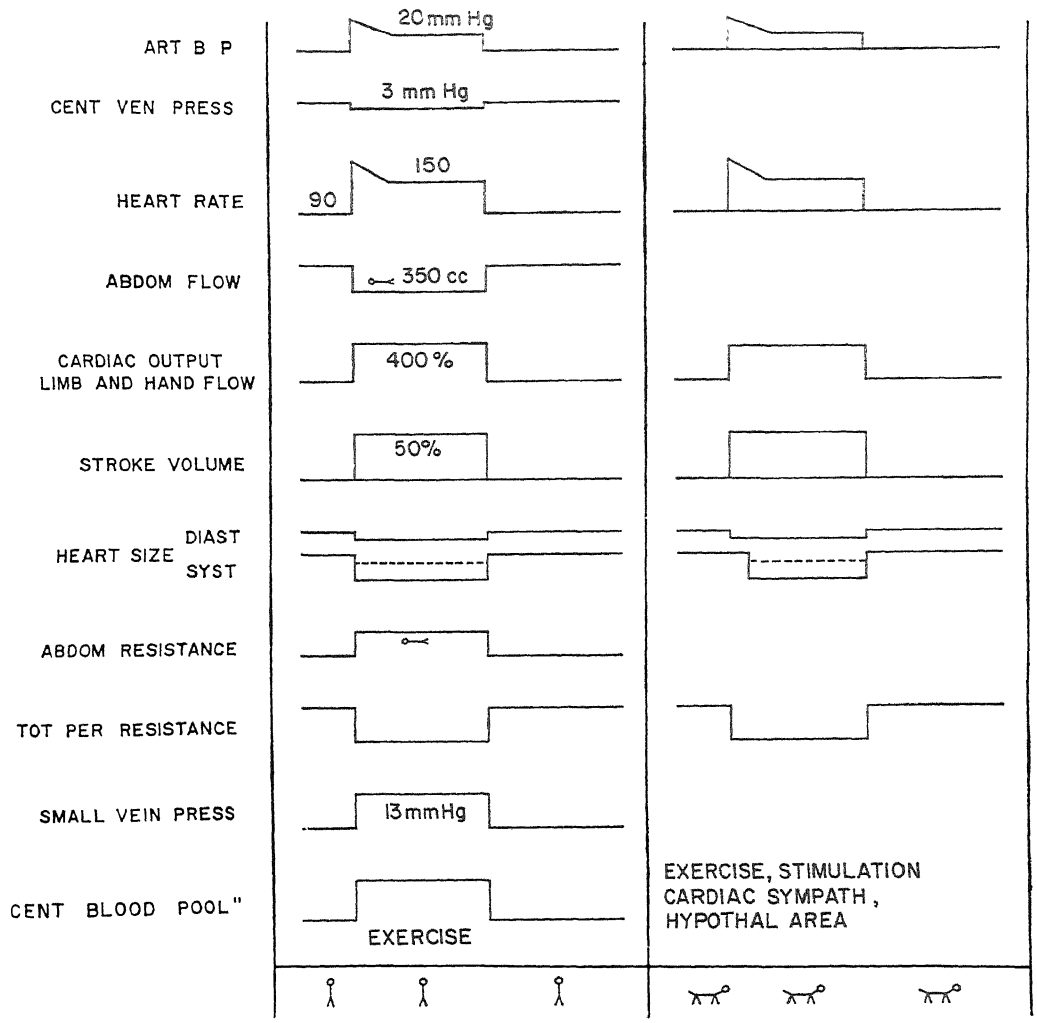


Fig 3 121 Left side hemodynamic responses in exercising man and right side hemodynamic responses in the unanesthetized dog to exercise to stimulation of the cardiac sympathetic fibers and of the hypothalamic area Dotted lines represent the less usual responses

number of different vascular beds in intact, conscious animals under a variety of physiological conditions Table 3 14 is presented showing data on distribution of flow and oxygen consumption in resting and exercising man, obtained by different types of measurement by different observers on different subjects (In some instances, values are extrapolated from animals) The values are certainly not absolute but the general pattern of distribution may be right The kidney blood flow of 400 ml /100 g of tissue is about six times greater than that of any other organ (nearest is the heart with a flow of 70 ml) The heart per unit of tissue, on the other hand, uses considerably more oxygen than any other organ Four regions—brain, heart, liver and intestine, and kidney, make up about 7% of body weight but receive 65 to 70% of the output of the heart and account for 60% of the total oxygen used at rest The skeletal muscles representing 50% of body weight receive only 20% of cardiac output at rest and account for only 15% of body oxygen usage, but during exercise they may account for 80 to 90% of the oxygen consumption and cardiac output The brain which is only 2% of body weight requires more than 20% of the total oxygen used by the body and during convulsions (observed in monkeys), this fraction may increase to 40% and the brain fraction of cardiac output from 14 to 20% Thus, the cardiovascular response is not limited to an adjustment of the cardiac pump, complicated as this may be Certainly, real changes occur in the distribution of the output of the pump

Coordination of Right and Left Heart Activity

As shown in Figure 3 122, the outputs of the left and right ventricles must be maintained in equilibrium if the blood volume and hemodynamic relations between the pulmonary and systemic circulations are to be preserved The cardiac factors which serve to interrelate right and left ventricular activity are

1) Frank-Starling Mechanism This serves to adjust each ventricular output to its inflow

2) Ventricular End-Diastolic Volume The ventricles do not empty completely with each beat A change in stroke volume induced by sudden alterations in sympathetic tone can occur from beat to beat through alterations in the ratio of stroke volume to end-diastolic volume In dogs, abrupt onset of exercise induces a simultaneous increase in both right and left ventricular stroke volumes (Franklin et al , 1962), presumably because of increased sympathetic activity Any disparity between venous returns of the two ventricles can thus be buffered by changes in end-diastolic volume

These factors, in turn, are aided by mechanisms which effect coordination of right and left heart volume output and accommodate temporary differences

1) Storage Function of the Pulmonary Circulation This permits moderate increases in pulmonary blood volume for a time Experimentally, rapid intravenous infusion of saline increases right ventricular output first The left ventricular output is increased similarly after a

Table 3 14
Distribution of cardiac output and oxygen usage

Region	Weight		Blood Flow			Oxygen Usage			O_2 Extraction	
	Kg	% Total	Liters/ min	% Total	Ml / 100 g /min	Ml / min	% Total	Ml / 100 g /min	Venous O_2 ml / 100 ml blood	AV O_2
Total	70	100	5 4	100		250	100		14 5	4 5
Brain	1 54	2 2	0 83	15	54	63	23	3 7	12 5	6 5
Convulsion			1 15	20	75					
Heart	0 33	0 5	0 22	4	70	23	9	7 0	7	12
Liver, intestines	2 86	4 0	1 54	29	54	55	20	1 95	15	4 0
Kidney	0 33	0 5	1 43	27	430	20	7	6 0	17	2 0
Skeletal muscle	34 0	50 0	0 92	17	2 7	55	20	0 16	12 5	6 5
Exercise			35 0	80	100	2700	90	7 9	3	16 0

lag time of at least three cardiac cycles During this period the blood volume in the pulmonary circuit increases temporarily

2) *Flow Transmission through the Pulmonary Vasculature* Right ventricular flow pulse is transmitted with little storage of volume in the lung circulation The pulmonary venous flow pulse is synchronized with the period of rapid filling of the left ventricle Therefore, the end-diastolic volume of the left ventricle can be directly influenced by right ventricular stroke output, which is transmitted to the left atrium in about 0 1 sec (Morkin et al , 1965) An increase in right ventricular stroke volume can cause an increase in left ventricular stroke volume at the next heart beat (Franklin et al , 1962) The transmission of this stroke flow is related to the stiffness (compliance) of the pulmonary vasculature Increased stiffness (sympathetic activity) increases flow wave transmission, thus aiding in ventricular coordination (see pp 3-22 and 3-210)

3) *Bronchial Circulation* There is free communication between the capillaries of the pulmonary vascular bed and the bronchial circulation The bronchial flow (part of left ventricular output) returns in part to the left atrium through the pulmonary veins, a smaller fraction returns to the right atrium via the azygos vein and superior vena cava Thus, a shunt is provided which could serve to equalize pressure and flow between the two circulatory systems The possible importance of this mechanism in health and disease has not been established

COORDINATION OF ATRIAL AND VENTRICULAR ACTIVITY

Atrioventricular conduction time decreases with the interval between heart beats so that the timing of atrial and ventricular contraction remains sequential An increase in rate of spread of excitation over ventricular musculature and shortening of the action potential duration accompany cardiac acceleration Thus, systole is shortened and the sequence of activation of the myocardium as a whole (i e , both ventricles) is maintained with changes in heart rate

THE HEART AND VESSELS AS A COORDINATED SYSTEM

The right and left sides of the heart are connected to their respective vascular systems at

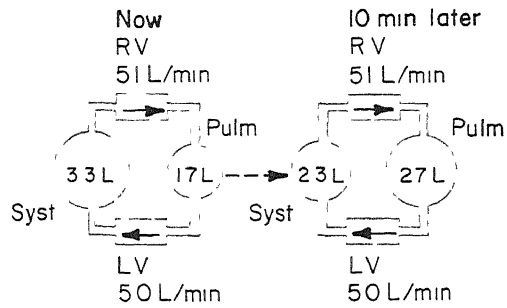


Fig 3 122 Diagram illustrating the consequences of a 2% difference in right (RV) over left (LV) ventricular outputs if mechanisms did not come into play to keep the mean outputs of the two ventricles the same Actually death would supervene before a shift of blood volume of the magnitude depicted could occur (Volumes and figure from Burton 1965)

both their input and output ends Cardiac output is adjusted to its venous return and to arterial pressure The arterial side is hemodynamically coupled with the heart only during systole, the venous side at all times Through these couplings, circulatory flow volume and pressure directly influence the intrinsic regulatory factors of the heart More remote influences from the vascular system are transmitted to the heart by nerve impulses and humoral agents (extrinsic regulation) Primary changes in cardiac output similarly affect vascular function Thus, the volumes, flows, resistances, and pressures in all parts of the circulation are capable of affecting function in all other regions of the system

Cardiac function curves (so-called Starling curves) relate factors governed in part by venous return (e g , mean right atrial pressure, end-diastolic ventricular volume, end-diastolic pressure) to cardiac output or some function of the ventricle which is a determinant of cardiac output (e g , stroke volume, ventricular pressure, rate of pressure development dP/dt, power, work) Figure 3 123 illustrates a group of such curves relating cardiac output to right atrial pressure Usually left ventricular output is determined, so that the curves describe a relationship between left ventricular function and right atrial pressure Such a simplified model of the circulation is shown in Figure 3 123*A* As can be seen from the normal curve in Figure 3 123*B*, when right atrial pressure increases, cardiac output increases When right atrial pressure falls sufficiently below zero (e g , −2 0 mm Hg), cardiac output is 0, at 0 0 mm Hg it is 5 liters/min and at 5 0 mm it is 12 5

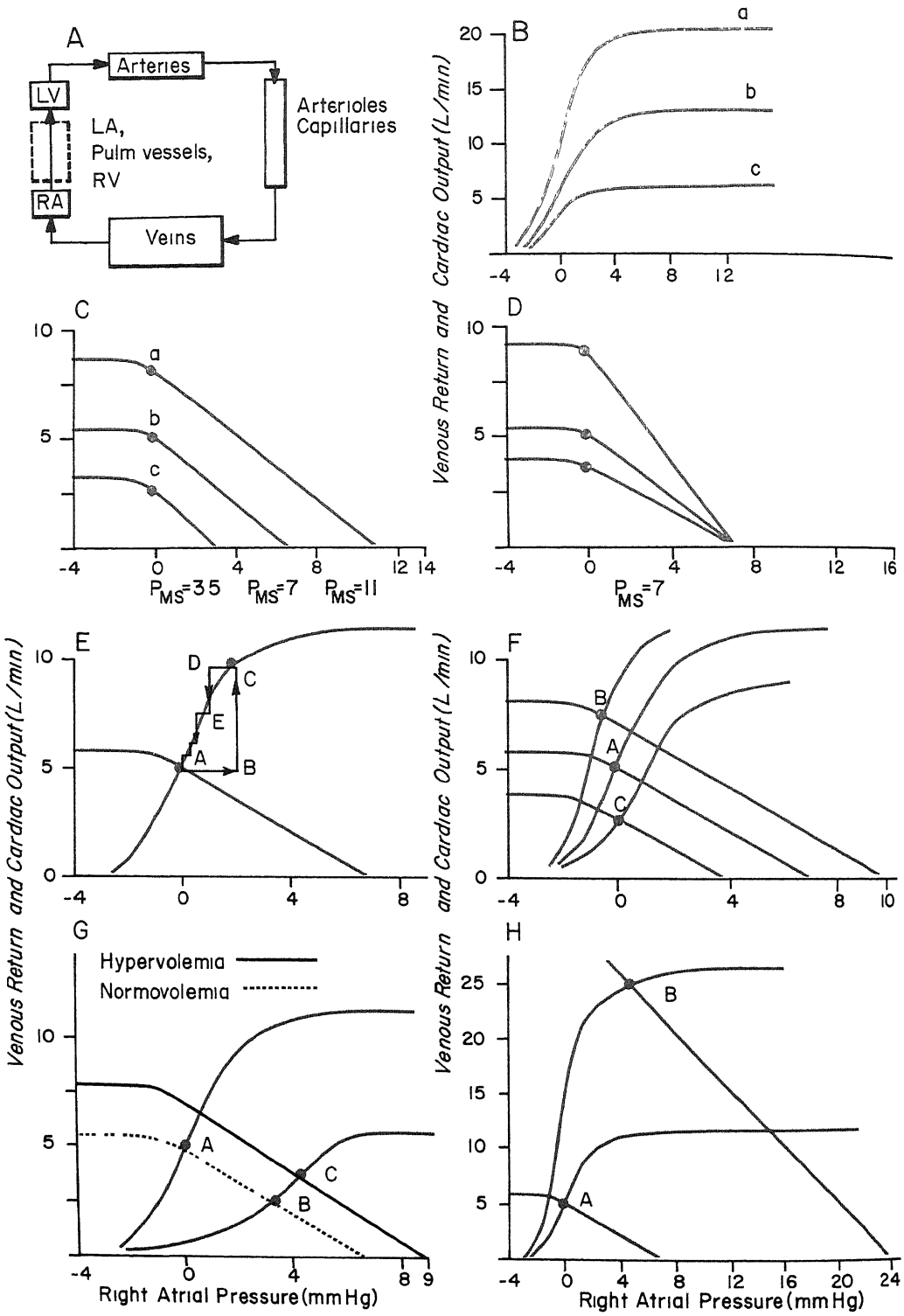


Fig 3 123 Cardiac and Systemic Function Curves *A* simplified model of the circulation Cardiac function curves may be plotted from determinations of right atrial pressure (RA) and left ventricular output (LV) The resultant curves characterize steady-state flow through the heart and lungs When the only parameters measured are LV output and

liters/min Lines *a* and *c* are for hyperdynamic (*a*) and hypodynamic (*c*) hearts The former represents the direction of change with, for example, sympathetic nerve stimulation or cardiac hypertrophy of lengthy physical training, the latter represents the direction of change associated with inhibition of sympathetic tone, vagal stimulation or myocardial disease

These curves permit an evaluation of the functional state of the heart based on the measurement of two factors e g , right atrial pressure and left ventricular output Any influence of the right ventricle, pulmonary circulation, or left atrium would alter the relationships between these two variables, but would not be distinguished from changes in left ventricular function by this analysis

Venous return or *systemic function curves* (Guyton, 1955, 1963) are employed to relate right atrial pressure and venous return (Fig 3 123*C*, *D*) The same two coordinates used in the Starling curve are here related right atrial pressure and volume flow But in this case flow is expressed as venous return rather than cardiac output Note, however, that now the relationship between flow and pressure is the inverse of that described by Starling curves In the Starling curve a rise of atrial pressure is associated with an increase in cardiac output In venous return curves, increasing atrial pressure decreases venous return The systemic function curves represent right atrial pressure (dependent variable) as a function of venous flow (independent variable) In the systemic function curves, contrary to convention, the dependent variable is plotted along the abscissa, whereas in the Starling curve convention is followed and the dependent variable (cardiac output) is plotted along the ordinate The fact that the relationship between pressure and flow is reversed in the two types of curves is not contradictory It simply means that the variables are interdependent and both relations operate simultaneously in the system (see pp 3-8 to 3-12) By considering flow first as the dependent variable (Starling curve) and then as the independent variable, (systemic function curve) the individual roles of factors determining right atrial pressure, venous return, and cardiac output are emphasized Starling curves deal with circulation through the heart and lungs, systemic function curves with circulation through the systemic vessels In the former, the heart (with the lungs and pulmonary circuit remaining to provide blood oxygenation) is allowed to respond functionally as the dependent variable to controlled alterations in pressure, flow, and resistance of the systemic vascular system which affect atrial pressure (independent variable) In the latter, the heart and lungs are replaced by a pump oxygenator (independent variable) and flow in the systemic vascular system becomes the dependent variable Figure 3 123*C* depicts a series of systemic function curves Note that in the normal curve, blood flow into the right atrium reaches zero at an atrial pressure of 7 mm Hg This is the expected pressure in all parts of the systemic circulation when equilibrium occurs following cessation of flow It is the mean filling pressure of the systemic circulation and is termed *mean systemic pressure* (P_{ms}) The value 7 mm Hg has been determined in dogs by stopping the heart, and within 2 to 3 sec pumping blood from the arterial to the venous side until the pressure is equalized (Guyton et al , 1954) The

---

RA pressure the possible effects of hemodynamic influences of the right ventricle pulmonary vessels and left atrium (dotted square) are ignored *B* cardiac function curves for the heart *a* Hyperdynamic cardiac state *b* Normal state *c* Hypodynamic cardiac state *C* Systemic function curves the effect of changing mean systemic pressure (P_{ms}) *a* P_{ms} of 11 mm Hg *b*, P_{ms} of 7 mm Hg *c* P_{ms} of 3 5 mm Hg *D* Systemic function curves the effect of changing resistance to venous return *E* Cardiac function (cardiac output) and systemic function (venous return) curves plotted on the same axes *A* the equilibrium point with coordinates of 5 liters/min cardiac output and venous return (ordinate) and 0 mm Hg right atrial pressure (abscissa) These coordinates represent stable values at which the system is operating When a perturbation occurs (such as a sudden shift of right atrial pressure from *A* to *B*) the coordinates for cardiac output reach a new point (*C*) In a stable system, a series of changes is initiated (*D E* etc) and the original equilibrium value (*A*) is regained *F* the effect of altering sympathetic tone *A* control cardiac function and systemic function curves *B* the effect of sympathetic nerve stimulation *C* The effect of sympathetic nerve inhibition (e g with spinal anesthesia) *G* Shifts in cardiac and systemic function curves in congestive heart failure *A* equilibrium point of control curves *B* new equilibrium point when the heart fails (decrease in cardiac output and increase in right atrial pressure) *C* equilibrium point established when the increase in blood volume which is expected in heart failure occurs Cardiac output is substantially increased owing to this shift in the equilibrium point *H* The effect of exercise on cardiac and systemic function curves *A* the control equilibrium point *B* the new equilibrium point when cardiac output and right atrial pressure are increased during strenuous exercise

P_{ms} is determined by the volume capacity of the vascular system and its degree of filling (blood volume) In the systemic function curves an increase of venous return is seen as atrial pressure falls (Fig 3 123*C*, *D*) until a plateau is reached This is the maximum venous return which is not exceeded as atrial pressure continues to fall The plateau is caused by progressive collapsing of the veins (at the thoracic entrance) which increases resistance to flow This automatically limits the rise in venous return with further fall in atrial pressure When the P_{ms} is increased (increased blood volume, decreased vascular capacitance) the curve is shifted upward and to the right, while with a fall in P_{ms} it shifts to the left and downward (Fig 3 123*C*) If the P_{ms} remains constant when vasodilatation occurs, venous return increases as resistance to flow decreases and the function curves rotate about the point of right atrial pressure which equals P_{ms} (Fig 3 123*D*)

Figure 3 123*E* illustrates the relationship between right atrial pressure, venous return, and cardiac output These two function curves cross at point *A*, the *equilibrium point* The coordinates of this point represent the cardiac output and mean right atrial pressure which characterize the state of the heart and circulatory system at the time the curves are derived

As long as the system is operating along these cardiac output and systemic function curves, it will tend to stabilize at this equilibrium point For example, a sudden increase in right atrial pressure from point *A* to *B*, all other factors remaining unchanged, would result in increased cardiac output (point *C*) This could be induced experimentally by withdrawing a given volume of blood from the arterial side and injecting it during ventricular systole into the venous side The increase in right atrial pressure would increase cardiac output (point *C*) while decreasing venous return The cardiac output would then exceed venous return and during the next four or five heart beats, blood from the heart and pulmonary circulation would be transferred into the systemic circulation Thus, atrial pressure would be reduced (points *D*, *E*, etc) as cardiac output decreases also The process would continue until the stable equilibrium point for the two function curves would be reached

With alteration in sympathetic tone, shifts occur as depicted in Figure 3 123*F*, increased sympathetic tone shifts the ventricular function curve to the left and the systemic function curve upward and to the right New coordinates are then established for the equilibrium point (*B*) representing an increase in cardiac output and venous return and decrease in mean atrial pressure Decreasing sympathetic tone has the opposite effect and the new equilibrium point (*C*) shows that flow into and out of the heart is less for a given right atrial pressure Marked reduction in cardiac output, as occurs in heart failure, shifts the cardiac output curve to the right and the equilibrium point lower on the systemic function curve (from *A* to *B*, Fig 3 123*G*) Even if cardiac function does not improve, fluid retention with hypervolumia increases P_{ms}, shifting the systemic function curve upward and to the right, establishing a new equilibrium point (*C*)

With exercise marked shifts in the function curves occur In Figure 3 123*H*, the steady state attained during strenuous exercise is depicted *A* represents the equilibrium point at a resting cardiac output of 5 liters/min and mean right atrial pressure of 0 0 mm Hg *B* is the equilibrium point attained with a fivefold increase in cardiac output and elevation of right atrial pressure to about 5 mm Hg Mean systemic pressure has increased to 24 mm Hg, the result of sympathetic contraction of capacitance vessels The systemic function curve slope is steeper than the resting curve because vasodilatation in the active muscles reduces resistance to flow through part of the systemic circuit The cardiac output curve is shifted to the left as the result of inotropic cardiac effects of sympathetic nerve activity and elevation of myocardial tension (Anrep effect, p 3-110), caused by increased arterial pressure (afterload) which accompany exercise The two curves equilibrate at *B*, a point on the cardiac output curve which is near its maximum

The foregoing type of analysis is only a rough approximation based on simplified assumptions regarding the vascular system and reasonable presumptions regarding the course of the cardiac output curves However, it serves to illustrate the complex interactions between cardiac output, arterial pressure, right atrial pressure, and venous return in cardiac regulation In this way the analyses have predictive value and agree with experimental observations (Guyton, 1963)

Cardiac Reserve

Present knowledge of the reserves available to the heart to cope with various stress impacts permits little more than their enumeration. The potential reserves are its heart rate, its systolic and diastolic volume, its blood flow, the oxygen content of the body mixed venous blood, the cardiac work, the changing efficiency of the heart, and possibly the stores of catecholamines.

The reserve in heart rate is about 130 to 140 beats/min., resulting in a maximum heart rate of 210. The reserve in stroke volume is 80 to 100 ml. since each ventricle normally contains about 150 to 170 ml., of which 60 to 70 ml. is ejected during systole. This figure is somewhat lower than the measured figure of 190 ml. for maximum stroke volume in Peruvian natives in the mountains, exercising to exhaustion on a treadmill. Attainment of such an increased stroke volume indicates an increased contractility of the myocardium. The reserve in blood flow approximates 30 liters/min., since maximum cardiac output is about 35 liters/min. in man working to exhaustion, and the resting cardiac output is about 5 liters. The potential reserve of oxygen within the body is about 900 ml. This is based upon the fact the body mixed venous blood contains 15 to 16 ml. oxygen/100 ml. blood, and the blood volume is 5500 ml. However, neither normal persons nor persons with heart disease utilize this oxygen to the full. With maximum exertion, the oxygen extraction approximates 14 to 16 volumes %. This is supplemented by the ability of the human body to take up a maximum of 3000 ml. oxygen/min., or an excess of about 2700 ml. over that in the basal state. The reserve work of the heart is at about 75 kg. m./min., since the maximum left ventricle work recorded is about 80 kg. m./min., and the resting left ventricle work approximates 5 to 6 kg. m./min.

6 THE CORONARY CIRCULATION

Anatomy

Two coronary arteries, the right and left, carry blood to the myocardium About 2 mm from its ostium, the left coronary artery divides into the left circumflex and the anterior descending branches The former runs in the atrioventricular (AV) groove to the left, ending in a posterior descending branch The anterior descending branch runs downward in the interventricular groove toward the apex Near its origin, septal branches are given off The right coronary descends in the right (AV) sulcus and ends posteriorly as several descending branches on the right and left ventricles The subdivisions of the main coronary rami descend superficially in the general direction of the apex and give off myocardial branches which course directly into the ventricular muscle At or near the apex, where the superficial muscles form a spiral and penetrate to form the innermost (subendocardial) layer of muscle, the terminal coronary arteries go along with the muscle to supply the inner layer of both ventricles and the papillary muscles Anatomical studies in both the dog and human heart indicate that vessel branching within the myocardium is related to muscle bundles, but numerous communications exist between the different muscle layers

As in any other vascular system, each of the two coronary arteries connects with its capillary bed, its superficial myocardial venous bed, and eventually with the right atrium (Fig 3 124) The epicardial branches of the coronary arteries and coronary veins also anastomose with each other and with extracardiac arteries and veins There are numerous arteriovenous shunts In addition to these pathways, the arterioles as well as the capillaries and superficial veins connect directly with both ventricular cavities by discrete, deep drainage channels, the arteriolu-minal, the arteriosinusoidal, and the Thebesian vessels which together comprise the deep coronary drainage circuit

In man, approximately 50% of hearts have the right coronary artery predominant, 30% have a balanced coronary circulation, and 20% have the left coronary artery predominant

The experimental animal most frequently employed in studying coronary vascular function is the dog despite certain anatomical differences In the dog, in contrast to man, there is a higher incidence of left coronary artery predominance and differences in detail of arterial and venous distributions (Detweiler et al , 1970)

MYOCARDIAL CAPILLARY CIRCULATION

The coronary microcirculation is organized as in other tissues True capillaries (endothelial tubes) arise from metarterioles whose walls have smooth muscle cells at irregular intervals Precapillary sphincters (or cuffs of smooth muscle) are located at the origin of the true capillaries The ratio of capillaries to myocardial fibers approaches 1 1 in the adult myocardium This "ratio" is the same for skeletal muscle However, in the latter fiber diameter is much greater—50 μ in skeletal muscle versus 20 μ in cardiac muscle—and there are about six times more capillaries per unit cross-section of myocardium

Coronary Circulation and Cardiac Metabolism

LEFT CORONARY ARTERY FLOW AND CHEMICAL COMPOSITION OF CORONARY SINUS BLOOD AS INDICES OF LEFT VENTRICULAR METABOLISM

It is not possible to quantitate accurately the metabolism of the right ventricle in the dog because its superficial anterior cardiac veins have many exits into the right atrium and because their contained blood is grossly contaminated by blood from the left coronary artery However, a large percentage of the drainage of the left myocardium is accessible In most in

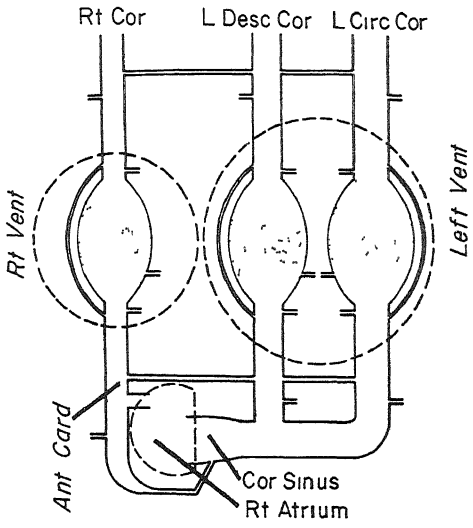


Fig 3 124 Schematic representation of the coronary circulation The areas enclosed within the circles depict the capillary beds collateral communications and AV shunts (Gregg and Sabiston 1956)

stances quantitative changes in the metabolism of the left ventricle can be obtained from measurement of the chemical composition of the coronary sinus blood together with measurement of the left coronary inflow This is because *1*) coronary sinus blood is almost entirely from the left coronary artery, *2*) by means of a special cannula which permits collection of all blood draining into the coronary sinus, the percentage of left coronary artery inflow recovered in the coronary sinus is quite high (80 to 90%) and reasonably constant during the induction of a variety of physiological variables and drug injections, *3*) the O_2 content of the venous blood from the left coronary artery not draining into the coronary sinus is generally only slightly less than that in the coronary sinus

Basal Data. In the resting state, the coronary data for dog and man agree With the left ventricular cardiac work index approximately 3 5 to 4 6 kg -m /min /m ^{2}, left coronary flow approximates 72 to 85 ml/100 g of left ventricle/min In anesthetized and unanesthetized dogs, under stress conditions, values as high as 600 ml /100 g of left ventricle/min have been recorded Although the heart can remove essentially all oxygen from the coronary blood in its passage through the myocardium, normally about two-thirds is extracted with an arteriovenous difference of 12 to 14 ml. and a coronary sinus value of 4 to 5 ml /100 ml of blood This extraction changes little with increased stress, signifying that the oxygen supply is well balanced with metabolic demands

Oxygen uptake per 100 g of left ventricle is 8 to 10 ml /min Oxygen consumption during systole averages about three times that in diastole

Myocardial Response to Anoxia. The coronary vascular bed is very reactive to the stimulus of anoxia, occlusion of a cornary artery for as short a time as 5 sec is followed by a temporary increase in blood flow Reactive hyperemia is considered to be the excess blood flow (over the control flow that normally would have occurred) which follows release of an arterial occlusion (Fig 3 125) The amount and duration of the reactive hyperemic blood flow increase with lengthening periods of left circumflex artery occlusion up to 120 sec , peak flows also increase but to a lesser degree The theoretical blood flow "debt" (control blood flow multiplied by the duration of occlusion) is usually greatly overpaid. Reactive hyperemia in skeletal muscle vascular beds shows these same characteristics except the blood flow "debt" is variably under or overpaid (see p 3–221) Coronary vasodilator agents do not consistently affect myocardial reactive hyperemia while Pitressin usually decreases the response

The *contracting* myocardium cannot withstand as long periods of arterial occlusion and oxygen deficit as *resting* skeletal muscle, and repays its oxygen "debt" with an increased blood flow but a decreased AV oxygen difference

DETERMINANTS OF CORONARY FLOW

Consideration of the mechanisms that control the blood supply to the myocardium involves certain difficulties not encountered in similar investigations in other organs of the body The myocardial wall of the left ventricle not only furnishes the pressure head for driving blood into the coronary arteries but also may either offer phasic resistance to coronary flow or actually aid flow by its muscular contraction around the coronary vascular bed Similarly, the right ventricle rapidly changes resistance to right coronary flow at the same time that left ventricular contraction presents blood to it under a pulsatile head of pressure However, coronary flow is related to the pressure difference (effective pressure) between the central coronary artery (identical to aortic pressure) and the right atrium divided by the sum of the

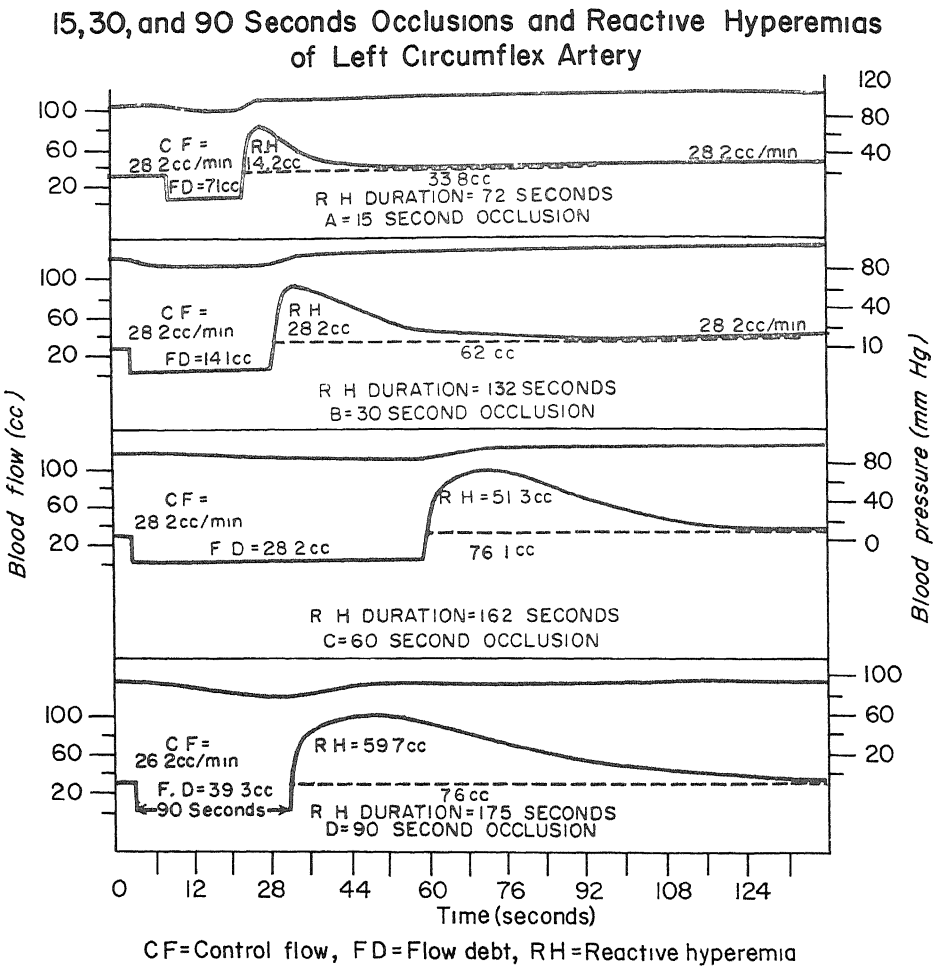


Fig 3 125 Myocardial response to anoxia Tracings redrawn from original experiment (Coffman and Gregg 1960)

viscous resistances to flow in the epicardial portion of the artery and in the peripheral coronary bed Viscous resistance to flow (aside from change in hematocrit) is mainly governed by the mean caliber of the coronary vascular bed The mean coronary diameter and, hence, flow are controlled by the effective intravessel pressure and by two peripheral mechanisms, i e , active changes in the state of the small mass of intravascular smooth muscle built into the coronary vessels, and the mechanical or passive effect on flow exerted during ventricular systole by the large muscle mass around the coronary vessels The peripheral mechanisms are regarded as more important in controlling coronary flow than the central coronary or aortic pressure which does not change greatly even in exercise It is estimated that doubling the aortic pressure might double the coronary flow, but doubling the average radius of the coronary bed might increase coronary flow about 16 times

Insight into the complexity of the integrating action of these three flow determinants has been obtained from recording of the peripheral coronary pressure and the phasic, or moment-to-moment changes, in coronary inflow and outflow in the epicardial arteries and veins (Fig 3 126) At the onset of isometric contraction of the left ventricle, there is an abrupt decrease in left coronary inflow (*solid line*) or even the appearance of backflow With the rise in aortic pressure, forward flow increases initially and rapidly, only to decrease to a new intermediate level in late systole With the onset of isometric relaxation, coronary flow increases significantly, peaking at early diastole and then declining progressively The velocity of coronary inflow differs somewhat from the

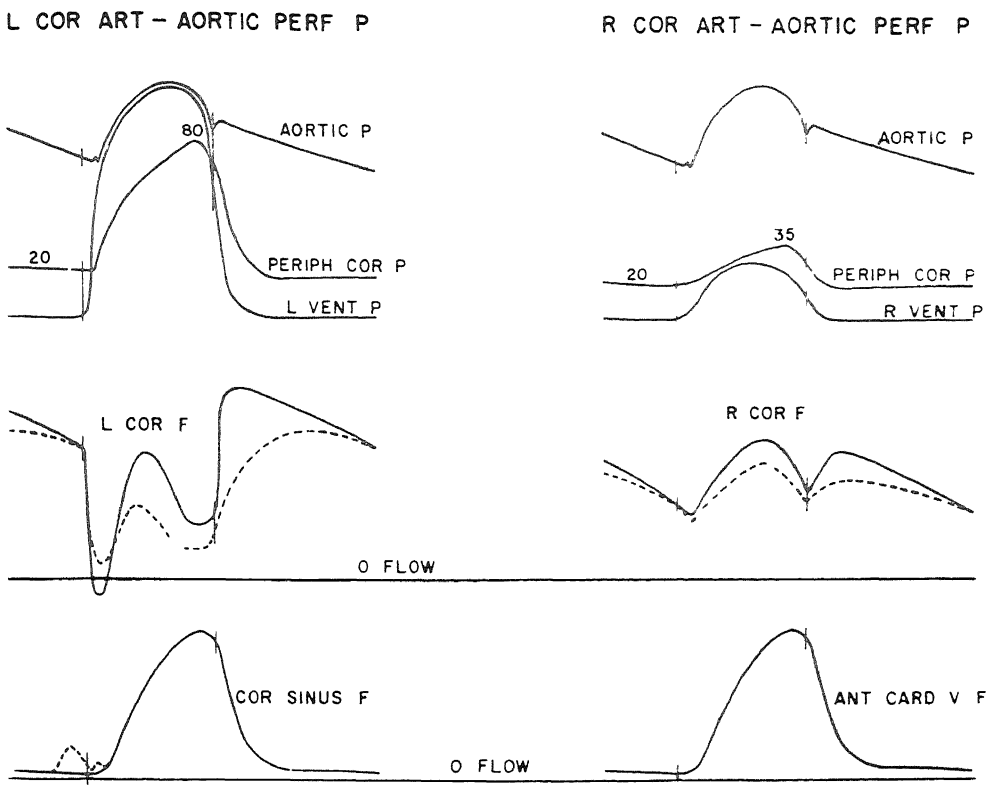


Fig 3 126 Series of curves relating phasic variations in left and right coronary inflow and coronary sinus-anterior cardiac vein outflow to aortic pressure ventricular pressure and peripheral coronary pressure Coronary inflow (*solid lines*) and anterior coronary vein flow obtained with orifice meter coronary sinus flow with pitot tube *Broken lines* estimated intramural velocity curves (Gregg 1959)

estimated intramyocardial flow (*dotted line*) The deficit during isometric contraction is caused by the compressing action of the myocardium on the coronary capillaries, forcing blood into the superficial vessels Early in the period of ejection, the flow excess is caused by the uptake of blood in the superficial coronary arteries, in diastole, the excess is caused by the uptake of blood to fill the previously compressed capillaries These demarcations of flow are much less obvious in the right coronary inflow pattern, which roughly resembles the prevailing aortic pressure curve. Thus, blood is flowing through the myocardium throughout the cardiac cycle, except possibly for a brief period in early systole in the left coronary artery In the left coronary artery, the systolic rate of flow is less than that during diastole, in the right coronary artery, the systolic rate equals or exceeds the diastolic In contrast to the left coronary inflow pattern, the flow curves in the coronary sinus and anterior cardiac veins rise and fall smoothly, with most flow occurring during systole and very little during diastole In unanesthetized dogs, using electromagnetic flowmeters, data have been obtained clarifying left coronary blood flow during systole and diastole In the resting dog, left coronary inflow during systole was 15 to 60% of diastolic inflow Under stress conditions (during excitement, chronic cardiac sympathetic nerve stimulation, or irreversible hemorrhagic shock), both the systolic and diastolic flow per heart beat increased by 300 to 400%, the ratio between them remaining the same, approaching unity, or even reversing, so that systolic flow surpassed diastolic flow Thus the strong contractions of the left ventricle, although impeding coronary flow, permit a large, but variable, systolic flow

The preceding account indicates that the coronary bed has a fluctuating resistance to coronary flow For the left coronary artery, the peripheral intracoronary pressure values at the height of systole and the low point of diastole approximate 80 and 20 mm Hg, respectively, and inflow is cut off at these pressure levels

when the left coronary artery is perfused under constant pressure In the right coronary artery, the contour and time relations of the peripheral coronary pressure curve are similar, but the values for systole and diastole and for the cutoff of flow are considerably lower

Elevation of right atrial pressure must increase coronary venous pressure and decrease the flow of blood in the coronary arteries provided no compensatory change occurs However, mild elevation of the pressure in the coronary veins draining the left coronary artery, by coronary sinus constriction, may decrease coronary flow and increase coronary arteriovenous oxygen difference only slightly Even marked constriction, or clamping, of the coronary sinus may cause only moderate reduction in arterial pressure and left coronary flow The smallness of the effect on coronary flow, despite elevation at times of the coronary venous pressure to near aortic systolic level, is presumably caused by the compensatory increased functioning of collateral venous communications with the anterior cardiac veins, in which flow increases considerably Similarly, right coronary inflow is not greatly reduced by clamping most of the anterior cardiac veins, presumably for the same reason

The relation of coronary perfusion pressure to coronary flow (*P/CF*) is generally such that the calculated coronary resistance decreases as the coronary flow rises This holds for almost any dynamic state of the coronary system (excised fibrillating or beating heart, heart in situ) and also for the pulmonary artery and certain peripheral vascular beds In the heart, it is impossible to say what fraction of flow change is caused by passive dilation from the perfusion pressure per se, since oxygen consumption per minute and per heart beat is elevated without any other observable systemic dynamic change Accordingly, if a physiological variable alters coronary pressure and flow by about the same amount and hence the ratio is unaltered, it is impossible to know whether the change in coronary flow is due to a change in effective head of pressure alone or whether the variable directly affected resistance However, a gross change in the vasomotor state of the bed occurs when coronary flow changes considerably with a constant pressure, or when flow and pressure change considerably in opposite directions.

Separation and quantitation of the determinants of coronary flow lying within the myocardial wall, i e , the intravascular and extravascular muscle, are of extreme importance To do this, continuous measurements are made in the open-chest dog while the left coronary artery is perfused with blood under a constant pressure, first it is done in the beating heart, and then during ventricular asystole induced by vagal stimulation or by disconnecting an external pacemaker which drives the ventricles, the latter manifesting complete atrioventricular heart block which had been surgically produced earlier By either means, the mechanical effects of ventricular contraction are largely removed Figure 3 127 illustrates that induction of ventricular asystole by vagal stimulation always increases immediately (within 1 second) left and right coronary inflow, in this instance from 56 to 78 and from 9 to 11 ml /min , respectively Thus, ventricular contraction acts to impede coronary flow through the ventricular wall The extent of the rise of flow is taken to represent the magnitude of the mechanical or passive factors limiting coronary flow The magnitude of this mechanical throttling effect on coronary flow during systole normally varies from 31 to 300% and averages about 50% The new flow level represents that state of coronary dilation related to the condition of the intrinsic smooth muscle of the coronary vessels at the prevailing coronary pressure The relative contribution of extravascular and intravascular re

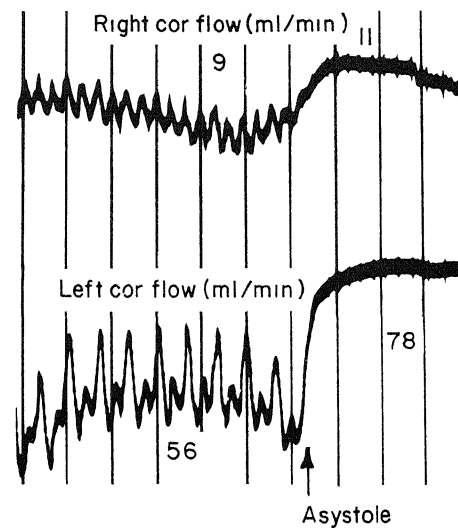


Fig 3 127 Response of left and right coronary flow to vagal induced asystole Note the prompt increase with the induction of asystole Time lines are 1 sec apart (Sabiston and Gregg 1957)

sistance to an increase of coronary flow has been tested under the different conditions of increasing heart rate, decreased arterial blood oxygen saturation, aortic constriction transfusion and drug injections In all instances, the major portion of a flow increase is through active dilation and not through reduction in extravascular resistance The largest reduction (40%) in extravascular resistance came from a decrease in arterial oxygen saturation

DETERMINANTS OF NORMAL CARDIAC METABOLISM

Metabolic Patterns in the Heart. The ability of the heart to do work depends basically on its biochemical activity leading to muscular contraction Cardiac muscle has been found to have basic chemical patterns similar to those of skeletal muscle The catabolism of fat, carbohydrate and protein produces free energy, about half of which is dissipated as heat and half is captured as phosphate bond energy which is used for muscle cell work and for various anabolic activities such as synthesis of glycogen, lipids, proteins and enzymes These catabolic and anabolic reactions proceed simultaneously under the influence of a complex system of enzymes, coenzymes (from the vitamin B complex), and hormones

Coronary sinus catheterization studies in man and dog have indicated that the heart is able to choose its fuel from a variety of foodstuffs These include mainly glucose, lactate, pyruvate, esterified and nonesterified fatty acids and, to a lesser extent, acetate, ketone bodies, and amino acids To determine their quantitative contribution to the energy production of the heart, i e , its oxygen consumption, measurements have been made of their cardiac extraction (coronary artery-coronary sinus difference), their total uptake [coronary flow times (coronary artery-coronary sinus difference of substance)], and the myocardial respiratory quotient (coronary sinus-arterial carbon dioxide difference) divided by (coronary artery-coronary sinus oxygen difference) Excellent correlation has been demonstrated between the myocardial respiratory quotient and the myocardial uptake of substance The extent to which each substrate contributes to the energy requirement of the heart in vivo is influenced by its concentration (above threshold) in arterial blood In addition, the state of nutrition of the organism markedly influences the kind of substrate used for energy production of the heart Under postprandial conditions, or after glucose infusion, myocardial metabolism is mainly glucose, lactate, and pyruvate, since its respiratory quotient approximates 0 9 with a high extraction of carbohydrate and a negligible uptake of amino acids Even the substitution of 5 to 10% oxygen for the normal 21% in the inspirated air does little to change carbohydrate uptake by the normal heart During overnight fasting, the heart derives much of its energy from fat, as indicated by a myocardial respiratory quotient of 0 8 with a low extraction and uptake of carbohydrate (see sect 3, chap 2, p 3-39) With prolonged fasting, the extraction coefficient for carbohydrate practically disappears, those for fatty acids and ketones are maximal, and the respiratory quotient is 0 7 With regard to the uptake of oxygen, the coronary A-V O_2 differences in man vary linearly with the arterial oxygen content through a range from mild anemia to marked polycythemia, so that the myocardial extraction coefficient (AV)/(A) is constant

Determinants of O_2 Usage of the Myocardium. Estimation of the metabolism of the myocardium in the absence of a heart beat, that is, during prolonged diastole, has been made in the vagal arrested heart as previously described under determinants of coronary flow Tht oxygen saturation of the arterial blood and coronary sinus blood is also measured continuously This permits left coronary arteriovenous oxygen differences as well as coronary inflow to be measured continuously first in the beating heart and then in the stopped heart until a new equilibrium is established, usually within 20 to 25 sec

Typical results are shown in Figure 3 128 As coronary inflow rises immediately with asystole from 70 to 90 ml /min , the oxygen saturation of blood in the coronary sinus also rises from a low control value of 44% to a very high value of 81%, thus greatly reducing the coronary arteriovenous oxygen difference The O_2 usage at rest is 35% of that in the beating heart Calculations in many experiments show that as the result of the combination of increased coronary flow and decreased coronary arteriovenous oxygen difference, the average oxygen usage per 100 g of left ventricle/min decreases from the average control level of 8 1 in the beating heart to 2 3 ml in the resting heart Individual values for the resting heart range from 1 0 to 4 0 ml This

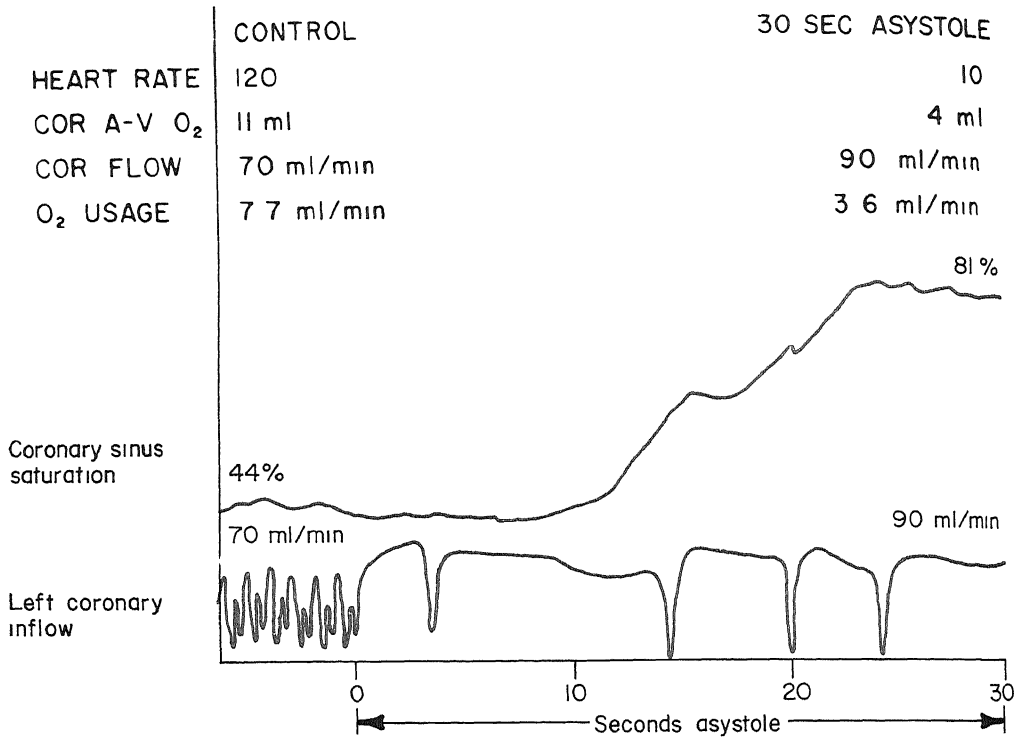


Fig 3 128 Reproduction of tracing from original record showing changes in left coronary inflow and coronary sinus oxygen saturation during prolonged ventricular asystole (Gregg 1955)

oxygen consumption in diastole is about one-third that in systole for an equivalent time period

Patterns of Response of the Coronary Circulation

The preceding gives some idea of the fractionation of the control of coronary flow and oxygen uptake within the myocardium during a cardiac cycle There are certain general patterns of hemodynamic responses of the heart and coronary circulation which follow primary changes in stress The information has been obtained from both humans and animals and has been assembled from many laboratories (Table 3 15)

METABOLITES

The coronary flow is greatly increased and the coronary arteriovenous oxygen difference decreased by some chemical constituents of the blood, while others produce no response or a decrease in coronary flow, in any case there is little or no change in systemic dynamics (blood pressure, cardiac work, heart rate) or myocardial oxygen consumption The coronary response to carbon dioxide and decreased pH is minimal, intracoronary injection of some intermediate metabolites will increase coronary flow and decrease oxygen extraction Although it has not been possible to demonstrate in the coronary sinus blood substances which would increase coronary blood flow, such metabolites could be metabolized or degraded rapidly and not be recoverable in venous blood Adenosine and adenine nucleotides are very active coronary vasodilators Although they are not found in coronary sinus blood, their derivatives, inosine and hypoxanthine, appear in measurable quantities during periods of myocardial hypoxia If adenosine is added to coronary arterial blood, only inosine and hypoxanthine are recoverable in the coronary sinus blood Thus it is possible that some metabolites may induce the vasodilation seen in myocardial hypoxia and yet not be recoverable in the coronary sinus blood

Autoregulation. As in other vascular beds, the coronary system tends to minimize the effects of alterations in perfusion pressure by

Table 3 15
Circulatory responses to stress states

Condition	Cardiac Work	Cardiac Output	Coronary Flow	Coronary Pressure / Coronary Flow	Oxygen Usage	Coronary AV O_2
↗Blood CO_2, ↘ blood pH[a]	→	→	→	→	→	→
Ischemia, mild anemia and anoxia, metabolites	→	→	↗	↘	→	↘
↗Heart rate, ↗cardiac output (transfusion) ↗Cardiac Work Aortic coarctation (mild),[b] thyrotoxicosis Cardiac sympathetic nerves, epinephrine and norepinephrine i v Severe anemia[c] and anoxia[c]	↗	↗	↗	↘	↗	→
Hypothermia, shock (hemorrhagic) Heart failure (extreme)[d]	↘	↘	↘		↘	→
Aortic coarctation (complete)	↘	↘	↗	↘	↗	→
Hypertensive cardiovascular disease	↗	→	→	↗	→	→

[a] Arrows merely indicate direction of response, not magnitude, i e ,→no change, ↗ increase or ↘decrease
[b] Cardiac output→or ↘
[c] Coronary AV O_2 ↘
[d] Coronary AV O_2 ↗

appropriate changes in resistance As illustrated in Figure 3 129, when the coronary perfusion pressure is increased or decreased, flow at first follows suit, but then resistance changes and flow returns toward the control level (see autoregulation (pp 3-6 and 3-164)

TACHYCARDIA

When the heart rate is increased considerably by electrical stimulation of the myocardium, aortic blood pressure, and minute cardiac output and work increase while the stroke volume and stroke work decrease Simultaneously, minute coronary flow and oxygen usage increase, coronary resistance decreases, oxygen extraction is unchanged but the coronary flow and oxygen consumption per beat decrease Since acceleration of the heart means proportionally greater time per beat and per minute in systole than in diastole and since in systole coronary flow is less than in diastole, it would be anticipated that increased heart rate per se should reduce coronary flow Since it does not, it must be that increased flow is due to arteriolar dilation from the increased metabolic activity The same trend of flow and oxygen usage per beat and per minute also occurs at the faster heart rate when minute cardiac work is held constant or when comparisons are made at the same stroke work This means that cardiac acceleration can augment the energy metabolism of the myocardium without manifestations of the extra energy as work

TRANSFUSION

Augmentation of ventricular load by increasing venous return and hence circulating blood volume through infusion has a clinical counterpart in the load placed upon the human heart by transfusion or by an aortacaval fistula The increasing coronary flow is partially explainable on a mechanical basis, since the slowing of the heart, which occurs under these circumstances, should increase coronary flow per beat and per minute by increasing diastolic time in which flow is greater However, active dilation presumably results from the increasing local chemometabolic activity associated with the increased cardiac work The coronary flow and oxygen are used economically, for the ratio

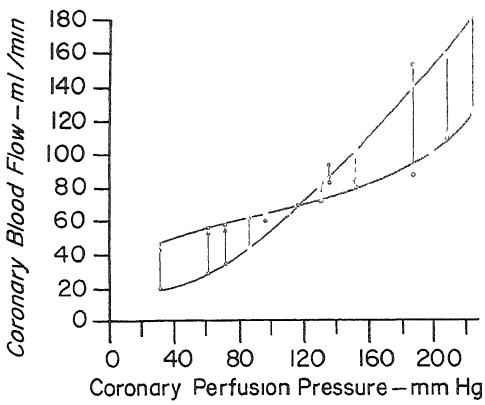


Fig 3 129 Pressure flow relationships in the coronary vascular bed The control level of coronary perfusion pressure and flow is at the point where the two lines cross Abrupt changes in perfusion pressure changed flow immediately to new levels (*closed circles*) which returned toward the control level (*open circle*) before becoming fixed at a new steady state (autoregulation of blood flow) (Redrawn from Berne and Levy 1967)

of cardiac work to oxygen consumption increases

AORTIC COARCTATION

With simulation of clinical coarctation by acute mechanical constriction (moderate) of the thoracic aorta just beyond the left subclavian artery, venous return to the heart by way of the inferior vena cava is decreased, but compensatory flow through various branches of the aortic arch may increase, so that cardiac output is maintained at the same level while the left ventricular workload is elevated With greater aortic constriction, the net cardiac output decreases, causing the cardiac work to decrease In either case the coronary dilation and increase flow arise in large part from *active* changes in the bore of the coronary bed related to the metabolic demands and in part passively from the increased blood pressure and moderately decreased heart rate The cardiac oxygen consumption is increased much more by this augmentation of pressure work than with an equal increase of volume work following transfusion No chronic studies of aortic coarctation have been made, because, owing to development of collateral circuits, the aorta may be first partially and then completely constricted at the arch without permanent development of hypertension proximal to the occlusion In human coarctation not much change has been reported in coronary flow and oxygen uptake per unit of muscle, but this might be expected because systemic pressure was only mildly elevated However, if true, the difference from the animal experiments might be explained by the fact that in these hearts, which are hypertrophied, there are fewer capillaries per unit of muscle to carry the oxygen and flow

HORMONES

Thyroid The myocardium participates in the increase in oxygen consumption characteristic of all body tissues in thyrotoxicosis This hypermetabolism is accompanied by an increase in coronary blood flow, a decrease in coronary vascular resistance, and an increase in oxygen consumption per minute and per beat Since there is an increase in oxygen usage per beat, cardiac oxygen utilization is presumably related not only to the increase in heart rate but to the general hypermetabolism of the myocardium as well

Epinephrine and Norepinephrine. The response of the systemic and coronary circulation to intracoronary artery injection of a minimal amount of epinephrine is similar to that indicated for stellate stimulation (see "Nervous Influences") Probably the coronary changes can be similarly explained With doses of 0 0002 mg or less, coronary flow may increase without any change in blood pressure or heart rate and with an increased coronary arteriovenous oxygen difference With the same or larger doses, as the systemic effects of the substance (increased aortic blood pressure, cardiac output, and changing heart rate) become evident, the coronary effects are exaggerated In those instances in which it has been tested, norepinephrine affects the coronary circulation similarly The circulatory responses to cigarette smoking or nicotine (i e , increased heart rate, systemic blood pressure, cardiac output, cardiac work, left coronary blood flow, and decreased coronary vascular resistance), parallel those with epinephrine or norepinephrine injection and are presumably related to the release of norepinephrine However myocardial oxygen consumption may increase, decrease or remain unchanged

Acetylcholine. This hormone, when infused into a coronary artery, increases coronary blood flow in the anesthetized dog If the dose is properly chosen, this response in the beating heart occurs without a significant change in blood pressure, heart rate, cardiac output, or

cardiac work, myocardial oxygen consumption per beat and per minute have been shown to increase However, when the hormone is fused intravenously into dogs with a constant heart rate and peripheral resistance, a decrease in myocardial oxygen consumption and ventricular contractility accompany the increased coronary blood flow The oxygen consumption of the myocardium for a given amount of cardiac work is not changed Although mechanical factors have been excluded, whether the effect of acetylcholine on coronary blood flow is directly on the intrinsic smooth muscle of coronary vessels or is induced through metabolic changes of the heart cannot be determined from the experimental data

Pitressin and Angiotensin. Of the hormones used clinically, pitressin and angiotensin increase coronary resistance to flow Coronary flow decreases during pitressin infusion, the reduction occurring throughout the cardiac cycle in the presence of an increased central coronary pressure There is a decrease in myocardial oxygen consumption and contractility, and in stroke work Although a reduced metabolic influence is presumably the cause of the decrease in coronary blood flow, a direct effect on the coronary vascular bed has not been ruled out Angiotensin produces an increase in coronary vascular resistance, its mode of action is unknown An initial cardiac depression is followed by a positive inotropic effect Blood pressure and cardiac work are increased

NEURAL INFLUENCES

Experimentally, vagal stimulation causes parasympathetic coronary vasodilatation independent of the chronotropic and inotropic actions of the vagus (Feigl, 1969) The extent to which parasympathetic vasodilatation contributes to coronary artery regulation has not been determined Stimulation of the cardiac fibers from a stellate ganglion in the anesthetized dog increases mean left coronary flow as the result of a decrease in systolic flow and a large increase in diastolic flow This is in contrast to the unanesthetized dog where the ratio of systolic to diastolic flow is variable Concurrently, left ventricular metabolism, cardiac output, and cardiac work increase while the systolic and diastolic sizes of the heart decrease The decreased ratio of pressure to flow indicates an increased mean coronary bore The major part of the dilation is explained by active changes in the coronary bed resulting from the release of an epinephrine-like substance (presumably norepinephrine) which directly enhances myocardial metabolism

Experimental evidence has been accumulated that alpha and beta adrenergic receptors are present in the coronary vascular bed Norepinephrine, which presumably acts only on alpha receptors, and epinephrine, which acts on alpha and beta receptors, produce a relaxation of helical strips of coronary vessels (200 to 500 μ in diameter), this relaxation can be prevented by beta adrenergic blockers (Zuberbuhler and Bohr, 1964) Also the decrease in coronary vascular resistance induced by intracoronary infusion of isoproterenol in the potassium arrested heart can be blocked with beta adrenergic blocking agents (Klocke and associates, 1964)

There is evidence that the alpha receptors mediate vasoconstriction, but the physiological importance of this has not been established (Hirche, 1967)

Carotid sinus (q v) reflex sympathetic control of the coronary circulation has been demonstrated (Feigl, 1968) and in the cat stimulation of sympathetic C fibers causes coronary vasodilatation (Brown, 1968)

ANEMIA

The coronary system participates actively in the circulatory adjustments to anemia For hemoglobin values of 10 g or more, the systemic circulation is essentially unaltered, and the compensation of the coronary system to the decreased oxygen-carrying capacity is similar to that with hypoxia, i e , an increased coronary flow without change in oxygen uptake When the hemoglobin values reach 6 to 8 g , the response of the systemic circulation is manifested by tachycardia, increased cardiac output, cardiac work, and a fall in peripheral resistance The coronary flow may now triple, coronary venous blood may contain less than 2 volumes % oxygen, the coronary arteriovenous oxygen difference may be 4 ml or less, and oxygen uptake is considerably increased The increase in coronary flow is related in part to the decreased blood viscosity and in larger part to the active dilation associated with myocardial hypoxia, which arises from the low hematocrit and from the increased metabolism Ultimately, myocardial failure will occur in severe anemia when the coronary vessels have approached maximal dilation and cannot further

compensate for the decreased oxygen-carrying capacity of the blood either by increased flow or by increased oxygen extraction In the presence of coronary stenosis associated with anemia, the effect of coronary arteriolar dilation in increasing coronary flow is minimized by the high fixed resistance of the stenotic artery, and myocardial depression and failure occur at lesser degrees of anemia

VALVULAR LESIONS

Acute elevation of right ventricular pressure by pulmonary artery constriction initially decreases right coronary flow, to be followed quickly by a sustained increase in systolic, diastolic, and mean flow in the right coronary artery and, to some extent, in the left coronary artery Upon release, right coronary flow temporarily increases still further During the sustained response, the systemic blood pressure can be fairly well maintained At the same time both the right ventricular work and metabolism increase, the former arising from the increased pulmonary artery pressure and cardiac output, the latter indicated by the increased right coronary flow and greater oxygen extraction from the coronary blood The coronary response to elevation of left ventricular pressure by aortic constriction central to the coronary ostia is similar to that with pulmonary artery stenosis In both instances, the sustained flow increase indicates a dominant influence of active coronary dilation over the increased mechanical flow-inhibiting effect of increasing extravascular compression which earlier was dominant However, in dogs with congenital subaortic stenosis, reduced coronary flows have been recorded (Pyle, Patterson, and Gregg, unpublished) In such instances the flow-inhibiting effect of extravascular compression is probably not compensated for by arteriolar dilatation This would account for the evidences of myocardial anoxia (myocardial necrosis and fibrosis, sclerosis of intramyocardial arteries, sudden death) in such hearts

HEART FAILURE

The heart shares with other types of muscle the characteristic that an optimum exists beyond which further stretching reduces the force of contraction and leads to myocardial failure In acute heart failure with progressive deterioration of the myocardium from pulmonary artery stenosis, the changes in coronary flow and oxygen usage per minute and per beat may be in the same direction as those described for the right myocardium during pulmonary artery constriction, but of lessor magnitude If the heart failure is severe enough, extravascular compression can become dominant over any active coronary dilation from metabolic proc esses, and coronary flow and oxygen usage may be normal or decrease, with the oxygen extrac tion at times reaching 90% The coronary circu lation in the heart failing with severe aortic stenosis undergoes similar changes When acute heart failure and chronic congestive failure simulating the human condition are induced by surgical complete heart block, changes in left coronary flow and ventricular oxygen consumption rather closely parallel alterations in the reduced left ventricular work In chronic left heart failure due to rheumatic, arteriosclerotic, and hypertensive heart disease, the coronary circulation apparently responds by a slight increase in oxygen usage through maintenance of the left coronary flow and increased coronary arteriovenous oxygen difference The increased oxygen usage corresponds with that indicated above for the right heart in one stage of failure Such hearts have considerable difficulty in transforming released energy into realizable work Studies of the coronary circulation in high-output failure from excessive transfusion or a chronic aortacaval fistula are not available

HYPOTHERMIA

Hypothermia is often used during open heart surgery The associated changes that occur which tend to reduce the coronary flow are a diminution in blood and muscle temperatures, cardiac output, heart rate, cardiac work, and oxygen usage by the heart, and an increased blood viscosity, and a greatly lengthened period of ventricular systole Opposing these factors are the relaxation of the major coronary vessels, which is known to occur with hypothermia, and dilation of the coronary bed caused by the hypotension per se As a result of these determinants, coronary flow is decreased at low temperature However, the percentage of reduction is greater in cardiac output than in coronary flow, which results in an increase in the coronary fraction of cardiac output at temperatures of 25° to 26° C There may be a decrease, an increase, or no change in peripheral resistance

in the coronary bed, whereas in the systemic bed an increase in peripheral resistance occurs Myocardial function appears to be adequate However, many hearts are apparently not too far from failure, because if total venous inflow occlusion (which decreases coronary flow close to zero) is now added to permit open cardiotomy, myocardial failure supervenes, as evidenced by elevation in mean right atrial pressure and postmortem findings This trend can be reversed by perfusion of the coronary system with small volumes of oxygenated blood

SHOCK

Standardized oligemic shock in dogs is characterized during the hypotensive phase by a decrease in cardiac output, systemic blood pressure, cardiac work, stroke volume, stroke work, and by an increase in heart rate and an adequate central venous pressure Coronary flow and coronary resistance are greatly decreased Coronary flow is generally greater and the resistance generally less than can be accounted for by a simple decline in arterial blood pressure At the same time, the oxygen uptake decreases, and the coronary arteriovenous oxygen difference is generally unchanged (The coronary response to sustained hypotension through spinal anesthesia or injection of procaine and Etamon is similar) With restoration to normal systemic blood pressure by reinfusion (intraarterial and intravenous routes are equally effective), coronary flow is greater and flow resistance is less than at an equivalent arterial blood pressure in the preshock state, and the augmented flow is maintained even if circulatory failure subsequently intervenes

The fact that early in the hypotensive phase ventricular end-diastolic pressure does not rise indicates that the functional capacity of the heart is adequate for the work performed At times after prolonged hypotension there may be evident cardiac dilation and elevated left and right atrial pressures, with the heart eventually proceeding to ventricular fibrillation or standstill, suggesting that myocardial depression or failure is partially responsible for the hemorrhagic shock syndrome Gross and microscopic evidence of myocardial injury appears in both reversible and irreversible shock Such myocardial depression could be caused by an insufficient coronary flow during either the hypotensive or the restoration periods The high coronary flow during the restoration period would seem to preclude an insufficient coronary flow during this phase as an adequate explanation During the hypotensive period, the actual coronary flow is greatly curtailed The problem is whether the associated sizable reduction in coronary resistance is sufficient to permit enough blood to reach the myocardium to prevent it from failing In some instances, at least, this loss of myocardial contractility is consequent upon an insufficient coronary flow, since the relation of atrial pressure to cardiac size can be reversed by increasing mildly with a pump left coronary flow without change in either the hypotension or blood volume

HYPERTENSIVE CARDIOVASCULAR DISEASE

An exception to the general picture of coronary compensation to increased systemic stress appears to be the response of the chronically hypertensive heart In essential hypertension with a normal cardiac output and elevated systemic blood pressure, the coronary flow and oxygen consumption per 100 g of ventricular muscle are unaltered while coronary resistance increases This increased resistance is also present in the renal and cerebral circulations Since these hearts are generally hypertrophied, *total* coronary flow and oxygen usage are increased This deviation (increased cardiac work with little change in coronary flow and oxygen usage per 100 g of ventricular muscle) is explainable if it is assumed that such hearts with known coronary artery disease have an increased amount of perfused fibrotic tissue

EXERCISE

Direct measurements of coronary flow have been made in normal unanesthetized dogs by applying chronically an appropriate electromagnetic flow meter (a somewhat modified Kolin type) directly to the circumflex branch of the left coronary artery Illustrated in Figure 3 130 are the changes in phasic and mean coronary flow. First, the dog stood on a treadmill, then ran for 3 min at 12 miles/hr at a 5% grade, and finally, stood at rest Although not measured, the blood pressure and cardiac output obviously rose in this dog as they have in other dogs A pulse pressure pattern has been sketched in for orientation purposes The control flow is 46 ml /min , with most of this oc-

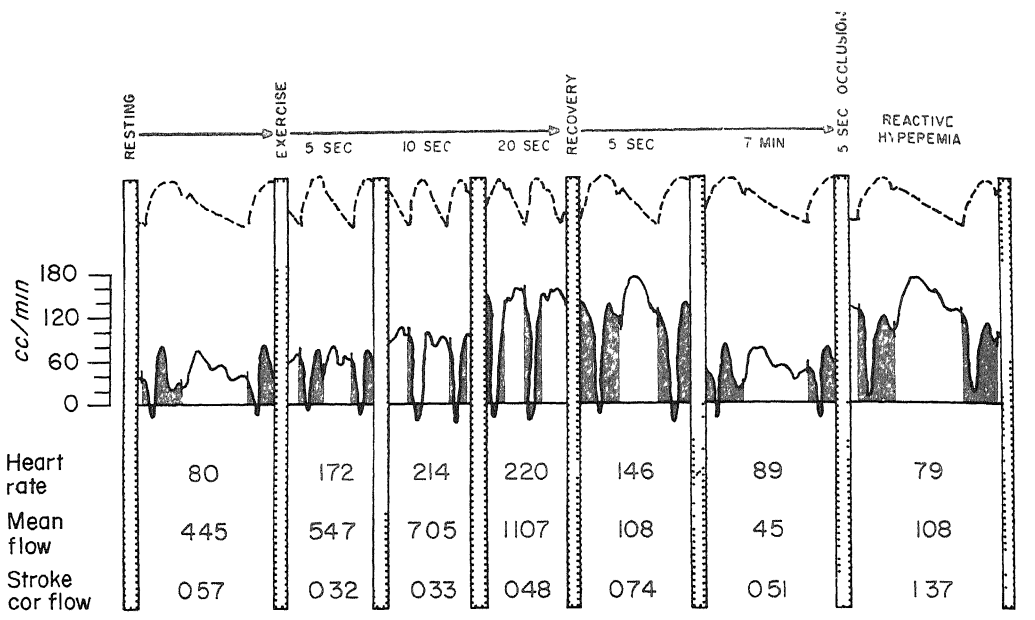


Fig 3 130 Response of the unanesthetized intact dog to exercise Curves are redrawn from the original experiment Phasic flow curves obtained with an electromagnetic flow meter monitoring an internal mammary-circumflex coronary anastomotic shunt *Upper curve* aortic pressure *Lower curve* phasic circumflex coronary flow (D E Gregg unpublished data)

curring during diastole Within 15 to 20 sec after the onset of exercise, the changes in the coronary flow response are complete and do not change further during the 3-min period of exercise The changes include an increase in heart rate from 80 to 220/min , a reduction in duration of both systole and diastole per beat and an increase in mean flow from 46 to 111 ml /min Calculation shows that the flow per heart beat is mildly decreased, and, hence, the mean flow increase is entirely on the basis of the increase in heart rate Within 5 sec after sudden stoppage of the treadmill, the patterns and flow have started to revert toward normal, the return to normal is completed within about 7 min The last record was taken so that a comparison could be made between the maximum diastolic rate of flow during exercise and that during reactive hyperemia It was taken shortly after the 7-min recovery period and immediately following a 5-sec occlusion by a pneumatic cuff of the coronary artery in which flow was being measured Despite the slower heart rate, maximum diastolic flow during reactive hyperemia is about the same as that during exercise

In other experiments, left coronary artery flow (electromagnetic flowmeter), cardiac output (a second flowmeter probe on the aorta), blood pressure (from a catheter in the aorta), and coronary sinus oxygen (from a catheter in the coronary sinus) have been monitored In *mild* to *moderate* exercise with little change in stroke cardiac output, with doubling of the heart rate, and with a mild blood pressure rise, coronary blood flow and myocardial oxygen usage *per heart beat* increased slightly, remained the same, or even decreased During induced excitement of the dog or stimulation of the cardiac sympathetic nerves, there was a massive dilation of the left coronary vascular bed, an increased heart rate with a greatly increased myocardial flow and oxygen usage per heart beat occurred Thus the coronary flow response to *mild* to *moderate* exercise is not similar to sympathetic stimulation of the heart in that stroke coronary flow and stroke oxygen usage are not always increased in the former The exercise response does resemble the effects resulting from a mechanical increase in heart rate by means of an artificial pacemaker (see also p 3–220)

SUMMARY

From the preceding discussion, it is apparent that when the primary change is an increase in

cardiac work and cardiac oxygen usage, concurrently, the coronary flow increases while coronary vascular resistance decreases In situations where the cardiac work and oxygen consumption decrease, so does the coronary inflow An exception to this general picture of coronary compensation to changing stress is that of complete aortic coarctation which may be characterized by a reduction in the output and work of the heart but an increased left coronary flow and oxygen usage A second exception is that of chronic hypertension in which the cardiac work is increased, but cardiac output, oxygen usage and left coronary flow are unaltered while coronary resistance to flow increases Thus, in most instances, coronary flow follows a primary change in cardiac stress However, coronary flow correlates best with oxygen usage This relationship is understandable since the oxygen consumption is derived from the product of the coronary flow and the coronary arteriovenous oxygen difference

Coronary Artery Collateral Circulation

The state of the human heart in the presence of naturally occurring coronary insufficiency or occlusion, or the state of the dog's heart in which coronary insufficiency has been experimentally induced, can be improved by augmentation of the naturally occurring coronary artery collateral circulation (by retrograde perfusion of the ischemic coronary bed with arterial blood) or by elevation of the ventricular fibrillation threshold In addition, in man, positive and subjective benefit from drugs or surgical procedures could arise through psychogenic effects which are not necessarily related to the heart

The measurement which is most important, least subject to error, and which has yielded most information, is that of the collateral flow This is determined by collecting the volume of blood that flows externally from a tube inserted into the peripheral end of an occluded coronary artery (retrograde flow) This is done acutely in the open-chest dog before, and sometime after, a variable has been chronically induced to improve collateral flow This is collateral flow before it passes through a capillary bed Collateral flow can also be measured after it has passed through a capillary bed and has appeared in the coronary sinus This can be done when extracardiac tissue with a vascular stalk has been previously placed on the heart to stimulate collateral development The coronary sinus flow is measured before and after clamping the potential extracardiac source of collateral blood Any difference in flow would represent that which has traversed the capillary bed

NATURAL RESPONSE OF THE CORONARY COLLATERAL CIRCULATION

The natural responses of the coronary system during experimental coronary artery constriction and occlusion (which presumably also occur in the heart of man) have been studied extensively Within 1 min after occlusion of a left coronary artery branch, the intracoronary pressure beyond this point drops to about 35/25 mm Hg, the ischemic muscle lengthens during systole of the left ventricle and useful function is lost When the peripheral end of this ligated coronary artery is permitted to bleed externally, collateral arterial blood appears immediately, averaging about 3 0 ml /min for about 50 g of potentially infarcted myocardium Probably not more than 2 4 ml of this blood (containing 0 5 ml of oxygen) would perfuse the myocardial bed if the collateral flow were not permitted to bleed externally This is because of the peripheral resistance existing beyond the point of occlusion This level of collateral flow and oxygen content is estimated to be up to 40% of that calculated as necessary to maintain indefinitely the viability of this myocardium (The oxygen uptake of 50 g of a heart with perfused coronary arteries at rest and doing no external work approximates 1 2 ml as compared to the immediately available collateral oxygen supply of 0 5 ml)

Hence, it is important to try to increase immediately this collateral or retrograde flow The level of retrograde flow does not naturally increase and cannot be made to increase for 8 to 10 hr by drugs or by any known physiological means such as increased heart rate, increased flow in the other coronary arteries or induction of hypoxia or hypoxemia in the other coronary arteries It can, however, be increased by passive elevation of arterial blood pressure (Fig 3 131) Why the collateral flow remains fixed, why the anastomoses function as a set of inert tubes, and why they do not participate in the vasodilatory response of the normal coronary bed is not known It contrasts with the rapid development of collaterals in other vascular beds.

This retrograde flow can be greatly reduced

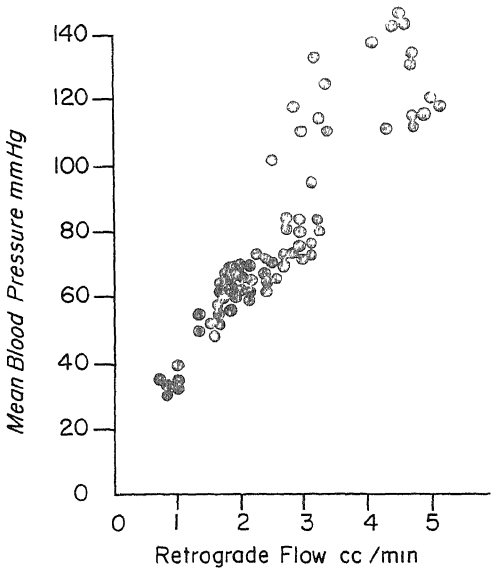


Fig 3 131 Relation of retrograde coronary flow to mean arterial perfusion pressure in dogs (Kattus and Gregg 1959)

by at least two conditions, namely, by excessive stretch of the nonischemic myocardium and by the presence of reactive hyperemia in the other coronary arterial branches One suspects that the reason for the decreased collateral flow in an occluded coronary artery branch while reactive hyperemia exists in the other coronaries is also the excessive stretch of the nonischemic myocardium Thus, the improvement that occurs in the human heart with drugs such as Neosynephrine in the presence of coronary insufficiency and infarction arises because of a good dynamic response in the non-ischemic but over-stretched myocardium This would increase the collateral flow by increasing the coronary perfusion pressure and by decreasing the size of the heart

Most hearts with an occluded coronary artery branch die within a number of hours If they survive the first few hours, then within 12 hr, for some completely unknown reason, collateral flow starts to rise, doubling within 2 days, and within 3 to 4 weeks it may approximate 40 to 100% of normal inflow into that coronary artery At the same time, the peripheral coronary pressure pulse beyond the point of ligation of the coronary artery branch may approximate, in pattern and ordinate values, the aortic pressure pattern simultaneously recorded

This is the usual course of events In other cases, stresses, some known but most unknown, prophylactically enhance the potential collateral circulation without the stimulus of coronary insufficiency or occlusion (Table 3 16) This enhancement is exemplified in man by the increase in the incidence of the injectable collateral bed in the presence of hypertrophy, valvular disease, cor pulmonale, anemia, and probably high altitude It is exemplified in animals by an increase in both the injectable and functional collaterals in the presence of high altitude and transfused anemia No good experimental evidence exists to indicate that physical exercise per se augments prophylactically the collateral flow as measured in a normal coronary artery immediately after its occlusion Even with a combination of moderate coronary artery constriction and exercise, the retrograde flow does not increase

DRUGS VS CORONARY COLLATERAL FUNCTION

Since the vascular bed beyond an occluded coronary artery is presumably maximally dilated by anoxia, it is unlikely that a drug could improve a heart by immediately increasing flow to this area Actually, the best experimental evidence offers no support for the belief that any drug penetrates the connecting collaterals or exerts any beneficial influence in the ischemic bed It has not been possible to increase

Table 3 16
Prophylactic augmentation of coronary artery collaterals by natural stimuli

	Known Stimuli	Incidence of Injectable Collaterals
		%
Man	None	9
	Hypertrophy	26
	Valvular disease	28
	Cor pulmonale	73
	Anemia	39
	Coronary constriction	11-63
	Coronary occlusion	90-100
	High altitude (man & pig)	↗
Pig	Anemia	100
Dog	Transfused anemia	↗
Dog	Mild coronary constriction and exercise	→

the flow of blood into this ischemic area within a short time after coronary artery occlusion, or to retain myocardial shortening by the use of drugs that are known to dilate the normal coronary bed, such as papaverine, nitroglycerine, epinephrine, aminophylline, Coramine, and khellin In chronic experiments, cortisone and ACTH do not promote collateral flow or reduce the size of infarcts The hypothesis that coronary occlusion reflexly decreases flow in other nonoccluded coronary arteries is negated by the experimental fact that flow in nonoccluded coronary arteries generally rises with coronary artery ligation The alleged favorable effect on survival of prophylactic and therapeutic drugs, such as papaverine and quinidine, is best explained by their known action in raising the fibrillation threshold and in reducing the excitability of the myocardium

NATURAL RESPONSE OF THE NORMAL BUT OVERSTRESSED PORTION OF THE MYOCARDIUM

Since the collateral flow does not increase for some hours, any early natural compensation must occur, not by improvement of the circulation in the affected area, but through enhanced action of the normal myocardium which is not involved Nonfunction of involved muscle following coronary artery occlusion reduces the total myocardial force available for raising intraventricular tension The immediate consequences, which lead to a hypodynamic ventricle, are a reduction in left ventricular systolic pressure, aortic pulse pressure, systolic and diastolic pressures, duration of systole, stroke volume and stroke work In this situation, left coronary inflow decreases considerably because of the mechanical reduction in size of its peripheral vascular bed However, within a few minutes, the nonischemic portion of the heart puts into operation compensatory mechanisms by means of which dynamic conditions are largely restored to normal, provided the nonischemic myocardium is in a good responsive condition Briefly, as the result of the lesser expulsion of blood with hypodynamic beats, the accumulating systolic remainder added to the oncoming blood stretches the viable muscle more and raises the ventricular end diastolic pressure or left atrial pressure This combination of increased pressure in and increased radius of the ventricular cavity causes the normal muscle to contract more vigorously, thereby restoring normal hemodynamics In this situation of increased cardiac work per unit of functioning myocardium, coronary flow, arteriovenous oxygen difference, and metabolism of the left ventricle increase The increase in oxygen uptake is equal to, and at times can be much more than that lost by the deletion of noncontractile muscle

However, not all hearts react as well because the viable portion of the myocardium may not respond to stretch, or may respond only temporarily This has been most thoroughly studied when the coronary obstruction has been induced by intracoronary injection of microspheres This leads to acute or progressive heart failure associated with a profound hypotension with a decreased cardiac output, and the clinical signs and symptoms of shock The clinical inference that this is due to the supervention of peripheral circulatory failure has not received heavy experimental support On the contrary, experimental work favors the view that, in this situation, circulatory failure not due to severe irregularity of the heart beat is due, successively, to defection of useful contractions in the ischemic area, a loss of contractile energy through expansion of the affected area, and failure of the still viable fractions to compensate adequately

ATTEMPTS TO IMPROVE CORONARY COLLATERAL CIRCULATION

The coronary arteries of the normal human heart are not "end arteries," but possess numerous collateral connections with adjacent vascular areas which can be increased by a variety of procedures (Gregg, 1969, Baroldi, 1969)

The state of the heart following coronary artery occlusion can be improved considerably by various physiological and surgical procedures The principles of the procedures used, together with the surgical manipulations, are shown in Table 3 17 These approaches include section of the cardiac sympathetic nerves, induction of myocardial hypoxia by various manipulations of the coronary venous system, production of chronic epicardial granuloma by mechanical and chemical means, application of extracardiac tissue to the heart, internal mammary artery ligation, sham operations, irradiation of the heart, coronary endarterectomy and coronary artery bypass

Animals. In the dog the coronary arteries function chiefly as end arteries and collateral

flow is normally negligible (Hammond et al , 1969) Hypoxemia (e g , coronary artery occlusion) activates interarterial anastomoses, increasing the effectiveness of collateral circulation (Blum et al , 1970)

Many of the procedures in Table 3 17 in the experimental animal are of positive benefit to the heart and give immediate or sustained protection against subsequent ligation of a major coronary artery ramus This is illustrated in the data obtained in many laboratories (Table 3 18) Ligation of a major ramus of the left coronary artery causes about a 70% mortality within the first 1 to 2 hr , and chronically there is considerable infarction When partial or complete occlusion of the coronary sinus precedes coronary artery ligation, or when a portion of the coronary bed is perfused in retrograde fashion by connecting the coronary sinus to an artery, the immediate mortality is reduced considerably With the exception of section of cardiac sympathetic fibers and internal mammary artery ligation, all other procedures listed—chronic coronary venous maneuvers, application of various

Table 3 17
Procedures for improving coronary function (man and animal)

Approaches	Tissue
Section cardiac sympathetic fibers	
Myocardial hypoxia	Coronary sinus constriction or ligation, aorta-coronary sinus shunt, pulmonary artery-left atrial shunt
Chronic pericardial granuloma via mechanical and chemical means	Talc, powdered bone, asbestos, mica, gelatin, sponge, silver nitrate, mechanical burrs
Application extra cardiac tissue to the heart	Spleen, omentum, fat, muscle, skin flap, intestine, lung, arterial implants
Internal mammary artery ligation	
Sham operation	
Coronary bypass and endarterectomy	Internal mammary and carotid arteries

Table 3 18
Physiological effects of prophylactic procedures on coronary collaterals (dog)

Experiment	Procedures prior to Ligation of Coronary Artery Ramus	Acute Ligation of Coronary Artery Ramus[a]			
		Mortality	Infarction	Injectable collaterals	Persistence of collaterals
		%			
Acute	None	70	—	—	—
Chronic	None	70	Gross	—	—
Chronic	Section cardiac sympathetic fibres	→	→	→	—
Acute	(Coronary sinus constriction or ligation) (Aorta-coronary sinus shunt)	↘	—	—	—
Chronic	(Coronary sinus constriction or ligation) (Aorta-coronary sinus shunt)	↘	↘	↗	Yes
Chronic	Irritants applied to heart· Talc, asbestos, mechanical abrasion, mica, phenol, silver nitrate	↘	↘	↗	Yes
Chronic	Mechanical abrasion, asbestos, coronary sinus reduced to 3 mm	↘	↘	↗	Yes
Chronic	Extracardiac tissue muscle, lungs, intestine, omentum, pedicle, skin flap, internal mammary artery	↘	↘	↗	—
Chronic	Ligation internal mammary artery	→	→	→	—
Chronic	Sham operation	↘	↘	↗	—
Chronic	Coronary bypass and coronary endarterectomy	—	—	—	—

[a] Directional arrows merely indicate direction, not magnitude, i e , →no change, ↗increase or ↘decrease

chemical and mechanical irritants, separately or in combination, and application of extracardiac tissue to the heart lead to a significant reduction in mortality and infarction

Man. All of the procedures listed in Table 3 17 designed to promote collateral development, including the sham operation, have been applied to the heart of man suffering from coronary artery disease All appear to increase to some extent the work and exercise tolerance and to decrease the cardiac pain These observations are not necessarily explained on the same basis as the improvement in the collateral circulation which follows such procedures in the dog This is because, in the dog, surgery precedes coronary artery ligation, whereas in the human, surgery follows coronary artery occlusion Surgery in the dog is designed to promote collaterals in the presence of a normal coronary circulation Surgery in the human is designed to promote collateral circulation after the coronary insufficiency has been naturally established In man, hypoxia, the greatest known vessel dilator, and a natural stimulus to collateral development, has already been working for many months To be effective, the surgical procedure must supply a stimulus to collateral development far more potent than hypoxia This has never been demonstrated in animals in which coronary insufficiency has previously been induced

The explanation of the results in man is not clear It may be that some subjects are protected and live longer because of the known experimental fact that handling the heart enhances the ventricular fibrillation threshold Some may be improved by procedures such as de-epicardialization which could obliterate the afferent pathways for pain However, results of the sham operation of Adams and of others, involving only a skin incision, strongly suggest that much of the positive benefit is on a psychogenic basis

Coronary endarterectomy, which has been applied to man, is on a sound physiological basis, and its purpose is entirely different from the preceding The surgeon directly re-establishes coronary flow through the original coronary artery by removing its atherosclerotic plug It does not require collateral development and should be effective provided there exists a gross coronary insufficiency of blood beyond the obstruction, provided the vessel remains patent and thrombi do not form, and provided there are no sizeable atherosclerotic lesions beyond the region of the occluded coronary artery Only a small group of patients have benefited from this type of operation and long term evaluation is not available Patients must be chosen carefully for this procedure with the above criteria in mind

Finally, bypass of a length of an occluded coronary artery by anastomosis of its peripheral patent end to a systemic artery has been recently applied therapeutically in man (Green et al , 1970) In dogs, a nonsuture anastomosis by intima to intima contact between the left coronary artery and the left internal mammary artery has been highly successful In almost all of the dogs, the anastomoses have been demonstrated to be patent as evidenced by gross observation, angiography, and measurement of coronary blood flow through the anastomosis up to the time of dog sacrifice (12 to 24 months after operation) End-to-side anastomoses of the internal mammary artery to 1 0 mm diameter segments of the anterior descending coronary artery is feasible in dogs, and results in long term patency and high flow Such a procedure was shown to relieve anginal symptoms and to increase exercise capacity in a small series of patients (Green et al 1970)

There is, however, a possible late complication to successful coronary endarterectomy or coronary bypass in man which must not be overlooked In the presence of such a large new blood supply, the existing collateral flow might disappear If another coronary occlusion subsequently occurs, the patient would be in difficulties because he had lost his collaterals

SUMMARY

From the preceding, it would appear that in the normal heart a moderate collateral circulation exists which functions immediately following a coronary artery occlusion Different prophylactic procedures are successful experimentally in preserving life either by compensating for a deficit in the supply of collateral oxygen or by preventing ventricular fibrillation This protection against occlusion of a coronary artery may be dependent upon a collateral development on the arterial side of the order of magnitude of that essential for viability Human coronary surgery which follows coronary insufficiency has no counterpart in animal experiments and these human coronary experiments cannot as yet be interpreted on a physiological basis

7 / MEASUREMENT OF BLOOD PRESSURE AND FLOW

Direct Measurement of Blood Pressure

In the direct approach, a catheter or needle is inserted into a vessel or heart cavity and connected to a hydrostatic column of fluid or to a special device Miniature transducers for implantation within the artery have been developed and employed for both hard wire recording and telemetering of blood pressure signals from free ranging animals (Van Citters and Franklin, 1969) Pressures thus obtained in almost all parts of the cardiovascular system of animal and man have been quantitated and studied Direct pressure measurements have value in determining certain aspects of function of the circulation They should not, however, as is pointed out elsewhere, be used as an index of change in flow through or volume of a blood vessel or heart cavity Whenever a change in volume is empirically deduced from a change in pressure, potentially active changes in caliber are ignored which are known to occur in every vessel except the capillaries

MEASUREMENT OF NONOSCILLATORY OR NEARLY NONOSCILLATORY PRESSURES

A vertical column of fluid with attached millimeter scale, and with some form of anticoagulant between it and the vessel whose pressure is being tapped, is the only device needed to determine the pressure in most intravascular pressure regions To measure the pressure, the column is filled from a side tube, and the column is then allowed to seek its own level by drainage into the vessel For measuring cerebrospinal fluid pressure, intravascular pressure in arterioles, capillaries and venules, and extravascular pressure in the tissue space in skin, muscle, kidney, subcutaneous tissue, etc , water or saline manometers are used in which little or no fluid enters or leaves the manometer This method yields higher values for tissue pressure, (+1 to +3 mm Hg) than those obtained with implanted perforated capsules, or in fluid spaces beneath the skin Interstitial fluid pressure measured by these methods is actually negative (−3 to −9 mm Hg) The existence of negative interstitial fluid pressure, although controversial, has its counterpart in negative pressures (−5 to −10 cm H_2O) measured in the peripheral cavity, joint spaces, and in subcutaneous spaces (Gauer et al , 1970) Obviously, the usual values employed in the Starling ratio for forces acting across the capillary wall must be modified to account for fluid exchanges where the interstitial pressure is negative (See sect 4, chap 12)

MEASUREMENT OF PRESSURE PULSES

Basically, all types of pressure apparatus involve measurement of the displacement of some type of elastic membrane which is moved by generation of pressure in a vessel or cavity In one group of manometers, an essentially nondistensible conduit (needle, lead tube, cannula, or catheter) is filled with fluid and connected to the membrane outside the body, this, in turn, is connected to a form of amplification and recording, in the other, a micromanometer membrane or its equivalent is mounted in a catheter tip at the pressure source, and the catheter serves only as a conduit for wires to the amplifying and recording assembly (Fig 3 132) The detection and transmission of the movements of the membrane can take place by mechanical, optical, or electrical means, or by a combination of these principles Owing to their technical imperfection, purely mechanical systems are now only of historical interest Pretty much in the same category are manometers which combine optical and electrical systems

Manometers Mounted in an Empty Catheter Tip. In this type of gauge, the pressure sensitive element is mounted in the catheter tip, the catheter contains no fluid, and merely

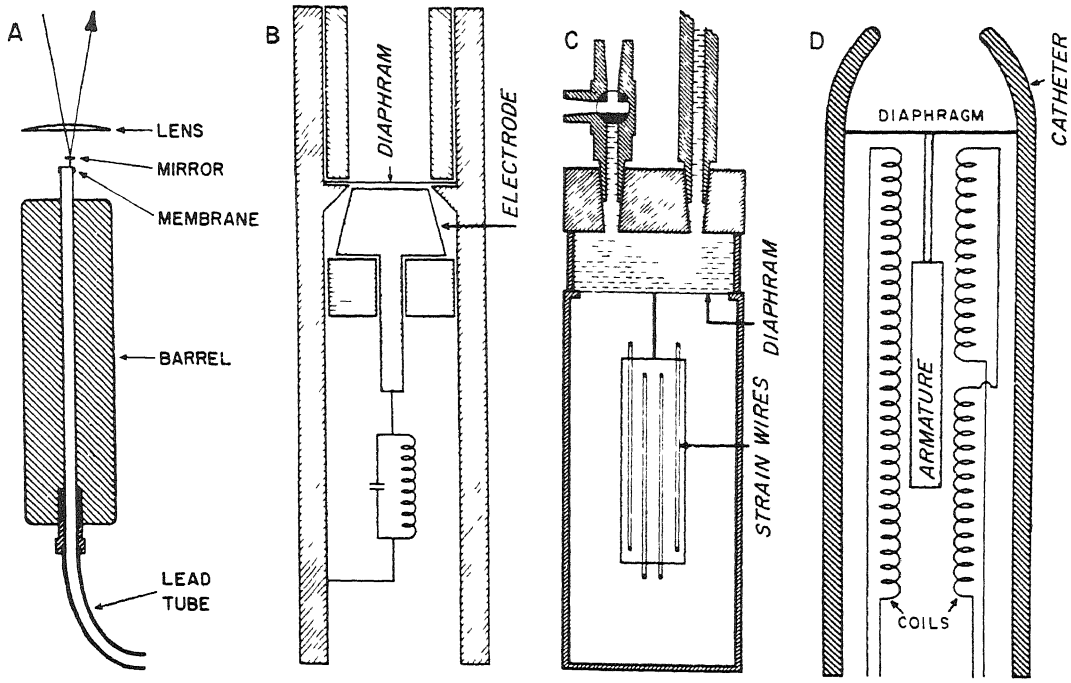


Fig 3 132 Externally recording pressure gauges *Left* principle of the classical optical method for determining biological pressures A flexible short nondistensible lead tube offering little damping effect connects to the pressure source *Center left* capacitance transducer *center right* strain gauge transducer and *right* catheter mounted inductance transducer

functions as a conduit for wire leading to an amplifier Inductance pressure gauges of this type are used in which the variable mutual inductance between a pair of coils can be altered by changing the position of an iron core within its magnetic field If the iron core connected to the center of an elastic membrane is mounted between two coils, movement of the membrane moves the core within the coils, changing their inductance Change in inductance, monitored by an appropirate bridge circuit indicates the extent of membrane displacement (Fig 3 132) This instrument was first devised by Wetterer (1954) and has been considerably modified and improved by others The strain gauge principle has also been applied to a micromanometer mounted in the end of a catheter In this, a strain wire wrapped around a metal reed with multiple longitudinal slits is stretched by the pressure source

Both the inductance and strain gauge type of micromanometer faithfully record the pressure pattern but are difficult and expensive to build Inductance gauges have been used in animals and, at times, in humans, but their use has been limited because it has not been possible to determine accurately the zero pressure with the Wetterer type Although the strain gauge type does not have this disadvantage and has been used with some patients, its initial expense is higher than the inductance type, and it has the added drawback that no way has been found to disconnect it from its catheter which wears out first

For precise quantitation, the intravascular pressure should be measured against the extravascular or tissue pressure which is around the vessel The terms, *effective*, *transmural*, and *net pressure*, have been introduced to indicate the pressure difference between the intravascular and extravascular pressure Precise determination of extravascular pressure, however, is unreliable

Indirect Blood Pressure Measurements

Although the blood pressure can be determined by inserting a hollow, wide-bore needle into an artery and connecting it with a suitable manometer, such a method is usually reserved for human cardiopulmonary diagnostic

work Convenience and safety demand that any method for general clinical use must be indirect The principle employed involves the balancing of air pressure against the pressure of the blood in the brachial artery The air pressure is estimated by means of a mercury or an aneroid manometer

The instrument used for this purpose includes an inflatable flat rubber bag covered by a nondistensible envelope of fabric, called the cuff The cavity of the bag is connected to both the manometer and a hand bulb or small pump by rubber tubing By this means the bag can be inflated to any desired pressure A small valve between the bulb or pump and the bag permits the escape of air, and the reduction of the pressure as required The uninflated rubber bag is wrapped snugly around the upper arm just above the elbow The bag is then inflated until the air pressure within it overcomes the arterial pressure and obliterates the arterial lumen The pressure is increased a little beyond this point and is then cautiously reduced again,[1] by the release of the valve, until the arterial pressure just overcomes the air pressure and blood escapes beyond the cuff into the peripheral section of the artery At this instant the pressure in the bag is read from the manometer Since the air pressure practically balances the systolic arterial pressure the manometer reading must indicate the value of the latter It is essential to the method that the manometer reading be taken at the instant when the blood escapes beneath the cuff

One of three methods, the *palpatory*, *oscillatory*, or *auscultatory*, may be employed to determine when the blood first escapes beneath the cuff In all three procedures the value for the lateral pressure in the brachial artery is obtained, whereas direct methods (which entail the introduction of a straight, unbranched cannula, or a large-bore needle into an artery) measure the end pressure The latter is greater because it includes not only the lateral pressure on the elastic arterial wall, but also that derived from the conversion of the kinetic energy of the moving blood column as it meets the obstruction, and the increment of pressure caused by the reflected wave from the obstruction

AUSCULTATORY METHOD

This procedure is the one generally employed clinically, It was introduced in 1905 by the Russian physician Korotkoff Certain sounds heard during auscultation over the brachial artery distal to the cuff are taken as the criteria for the systolic and diastolic pressures Under ordinary circumstances if a stethoscope is placed over the brachial or any other artery, no sound can be heard as the flow of blood through the channels is inaudible If, however, the artery is compressed by the manometer cuff so as to completely arrest the flow of blood, a sharp light tapping sound, in rhythm with the heart beat, will be heard when the pressure in the cuff is released and falls just sufficiently to permit the arterial lumen to open As the pressure in the cuff is progressively lowered the sound undergoes a series of changes in quality and intensity

Four phases of the sound, each having its distinctive character, may be heard in succession in the normal individual, as the pressure is gradually reduced from about 120 to 80 mm Hg or less These are described below with the average pressures at which they are normally heard (Fig 3 133)

Sounds of Korotkoff Phase I Sudden appearance of a clear, but often faint, tapping sound growing louder during the succeeding 10 to 14 mm Hg fall in pressure

Phase II The sound takes on a murmuring quality during the next 15 to 20 mm fall in pressure

Phase III Sound changes little in quality but becomes clearer and louder during the next 5 to 7 mm fall in pressure

Phase IV Muffled quality lasting throughout the next 5 to 6 mm Hg fall After this all sound disappears

Phase V Point at which sounds disappear

The muffling and disappearance are commonly referred to as the fourth and fifth points If there is a detectable difference between these, it is recommended (Kirkendall et al, 1967) that they be recorded e g , 142/82/78 mm Hg (142 = systolic, 82 = fourth phase, 78 = fifth phase) If the fourth and fifth points

[1] Should deflation be carried out too rapidly the mercurial type of clinical manometer shows a considerable lag, due to the time required for the air above the mercury column to reach atmospheric pressure With rapid deflation, particularly when the heart rate is slow, the actual point at which arterial pressure overcomes cuff pressure may be missed The rate of deflation should not exceed from 2 to 3 mm Hg/sec

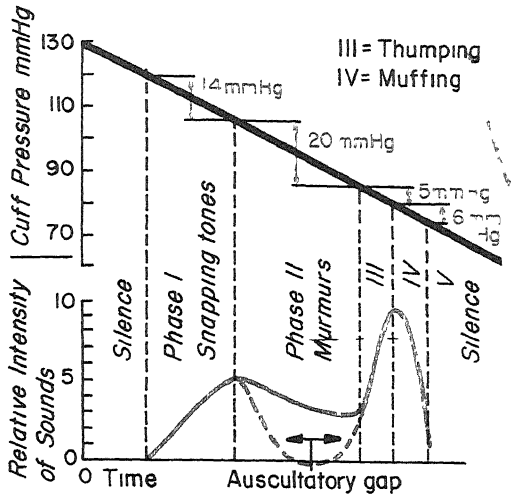


Fig 3 133 Characteristics of the auscultatory method of measuring blood pressure (Redrawn from Geddes 1970)

are identical, this should be recorded e g, 144/76/76 Muffling (Phase IV) occurs at pressures 7 to 10 mm Hg higher than direct intra-arterial diastolic pressures Disappearance of sounds (Phase V) occurs close to intra-arterial diastolic pressures in adults at resting heart rate With increase in flow as with exercise, it falls far below true diastolic pressure because the opportunity for intra-arterial turbulence is greater Under such circumstances Phase IV and V may be separated by 40 mm Hg or more, with Phase IV being closest to true diastolic pressure In children, in whom flow velocity is often high, Phase IV has been found the more reliable index of diastolic pressure

Auscultatory Gap. As shown by the sound intensity curve in Figure 3 133, the Korotkoff sounds may temporarily disappear during the latter part and Phase I and Phase II This gap may cover a range of 40 mm Hg and cause marked underestimation of systolic pressure if the cuff pressure is increased only to the level at which the sounds are absent This may be avoided by checking the palpatory systolic pressure first, when the palpatory systolic pressure is highest it should be accepted as the reliable determination Venous congestion associated with slow cuff deflation or rapidly repeated occlusions increases the occurrence of auscultatory gap Thus rapid cuff deflation (2 to 3 mm Hg/sec), allowing 2 to 3 min to elapse between determinations, and raising the arm to promote venous drainage, reduce or eliminate auscultatory gap

Arterial blood pressure cannot be measured with precision by means of sphygmomanometers The method is attended by inaccuracies caused by variations in heart rhythm, the presence of normal respiration, and resistance of the tissues of the part To compensate for tissue resistance, a rule of thumb is that the inflatable arm bag should be approximately 20% wider than the arm diameter, or approximately 12 cm for adults, 8 cm for children less than 8 years, 5 cm for children less than 4 years, 2½ cm for children less than 1 year For the adult thigh the bag width should be approximately 18 cm and such bags are useful also on the upper arm of obese individuals Despite use of such cuffs, a mean error of ±8 mm Hg may be expected in individual readings of systolic and diastolic pressures In general, the auscultatory method tends to underestimate systolic pressure and overestimate diastolic pressure In the arms of people with much fat tissue, the error is further exaggerated, in arrhythmias, deep breathing, or abnormalities of the arterial wall, the errors are still larger Increases in the resistance of the arterial wall, as a result of sclerotic changes or simple hypertonus of the muscular coat, give readings that are too high Repeated compression and decompression just before the actual determination is made may remove any spasm of the arterial wall sufficiently to reduce this source of inaccuracy Though lower readings as a rule are not obtained in a normal individual by repeated readings, in others with apparently normal arteries, the reading obtained after the third or fourth trial may, as a result of the reduction in tone of the vessel wall, be lower than the initial observation by several millimeters

Determination of the blood pressure in the human thigh by the auscultatory method gives incorrectly high values and has led to the misconception that leg blood pressure is higher than arm blood pressure Simultaneous direct intra-arterial measurements in the brachial and femoral arteries have shown the two pressure curves to be similar in all respects (Pascarelli and Bertrand, 1964)

OTHER INDIRECT METHODS

A variety of other methods based on the use of an occluding cuff have been devised in which

various sensors are employed for detecting sound, pulsation, or blood flow below the cuff

Flush Method In this method, cutaneous blood is forced out of the extremity below the cuff, the cuff is inflated above systolic pressure and slowly (6 to 7 mm Hg/sec) deflated Systolic pressure is indicated when the pale, bloodless skin flushes Applied to the adult finger the method gives systolic values below the auscultatory method and is unreliable if digital vasoconstriction is induced by emotional or temperature changes The method has found practical application in newborn infants applied to the hand or foot

Pulse Detectors. Various types of pressure transducers have been applied below the cuff to detect the earliest arrival of pulsation as the indicator for systolic pressure The size of the signal obtained varies as pulsation increases and these characteristic variations have been used to estimate (in general unreliably) mean or diastolic pressure

Microphones. Korotkoff sounds may be amplified by means of piezoelectric microphones mounted within or below the cuff The electrical signal obtained may be amplified to increase audibility or used to actuate a voltmeter or other visual signals such as a flashing light

Ultra-Sound. Ultra-sound may be used to detect arterial wall movements as pressure is decreased in the blood pressure cuff Ultrasound transmitting and receiving crystals are placed over the artery beneath the cuff The doppler signal received identifies the opening and closing of the artery At cuff pressure just below systolic level opening occurs only at the peak of pressure, followed rapidly by closure The time interval between these two signals increases as cuff pressure declines until it is low enough that the vessel remains open during the pulse

INDIRECT MEASUREMENT IN ANIMALS

Indirect blood pressure determinations in animals are made with adaptations of the methods described for man An additional extremity, the tail, is available in many species and exteriorized arterial loops may be employed for application of the occluding cuff The precision and accuracy of these methods decreases with decreasing size of the subject owing to difficulty of pulse and sound detection and vasomotor changes in the muscular distributing arteries to the extremities under various conditions

Equipment and methods for radiotelemetry of blood pressure and flow have been developed and employed in free ranging terrestrial and aquatic animals (Van Citters and Franklin, 1969a, b)

Measurement of Volume Flow of Blood

The most important measurement in the circulation is that of the volume flow of blood through the aorta or through any vital region or organ The blood flow is most sensitive to a change in dilation in a vascular bed If the blood pressure is doubled, the blood flow might also double but doubling the mean bore of a vascular bed might increase the blood flow by 16 times In the past, advancement of flow methodology has been slow and limited in large measure to devices applicable under highly abnormal conditions In use, the animals have been exposed to so much insult by anesthesia, surgery and trauma that the measurements do not necessarily have resemblance to normal values, i e , the relative importance of factors regulating flow can be quite different than in the normal state

VISUALIZATION (HUMAN OR ANIMAL)

Direct microscopic *in vivo* visualization of the components of the microcirculation (terminal arteries, arterioles, capillaries, venules, and very small veins) offers a means for studying qualitatively the peripheral circulation and control of blood flow in various regions When capillaries, arterioles or any of the above vessels are observed to decrease in diameter, they are assumed to have constricted and thus to decrease flow Despite this, such observations have contributed greatly to our knowledge of the circulation Direct microscopic and high speed photography have been used in the frog skin, bat wing, rat mesoappendix, dog omentum, lungs of the open-chest dog, transparent viewing chamber of the rabbit's ear, bulbar conjunctiva and skin of humans, and pial vessels of the brain through a transparent viewing chamber Change in small terminal arteries (and possibly arterioles) has been visualized in the coronary circulation of the heart of the closed-chest dog and in humans with

cinéphotography following injection of contrast medium directly into the coronary arteries

VENOUS DRAINAGE

One of the oldest, simplest and most satisfactory means of determining blood flow is to collect and measure the outflow from a cannulated vein draining a particular organ The rate of flow can be read from a graduate, or from the change in weight of a collecting container on a scale For very low rates of flow, drop recorders are used in which each drop shorts a pair of contacts in an electrical system or interrupts a photocell light beam

PERFUSION SYSTEMS

For many investigations and certain surgical procedures, it is desirable to have controlled perfusion of an organ, body region, or the whole body at constant flow rates and/or at constant perfusion pressures, which are the same or different from the prevailing aortic pressure Pump systems have been used to perfuse isolated organs and to perfuse, in the unanesthetized as well as the anesthetized animal, the vessels of an extremity the brain, liver, kidney, and the coronary vessels of the heart Mechanical pumps serve also at temporary replacement for the left ventricle, right ventricle, and the whole heart, i e , the whole body including the coronary arteries is perfused with a pump Their use for bypass of the right heart, left heart and whole heart is shown in principle in Figure 3 134 *J*, *K*, and *L*

Mechanical Circulatory Assistance (Cardiac Assist Devices). Mechanical assistance to the circulation is the application of any energy-producing device to aid blood flow Methodological complexity extends from simple manual cardiac massage, temporary partial or complete cardiac bypass, through to total cardiac replacement with an implantable artificial heart The chief techniques which have been employed are listed in Table 3 19 and illustrated in Figure 3 134

Cardiac massage, external (Fig 3 134 *A*) or internal (Fig 3 134 *B*), is applicable for assistance for periods up to several hours External manual massage by compressing the heart between the sternum and spine is an effective emergency procedure used to maintain circula-

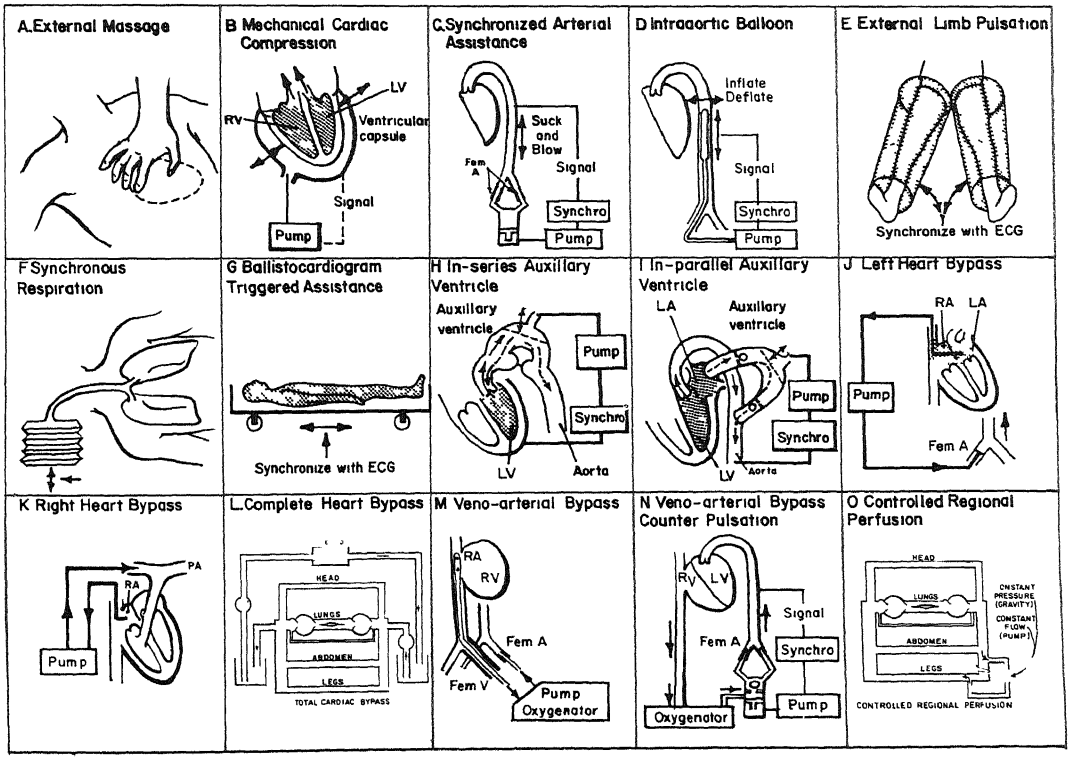


Fig 3 134 Schematic representation of circulatory assistance devices See text

Table 3 19
Methods of mechanical circulatory assistance

Cardiac massage
External manual compression of the thorax
Internal manual or mechanical cardiac compression
"Counterpulsation"
Synchronized arterial assistance
Intra-aortic balloon
External limb pulsation
Synchronous respiration
Ballistocardiogram-triggered assistance
("In-series" auxiliary ventricle)
Auxiliary ventricle
"In-series" (counterpulsation)
"In-parallel"
Left heart bypass
Right heart bypass
Veno-arterial bypass
With oxygenation
With oxygenation and counterpulsation
Implantable cardiac substitute

tion until the cause of cardiac arrest can be corrected or more effective measures applied

The object of *counterpulsation* is to reduce the pressure against which the ventricle must pump during systole (afterload) and to improve coronary flow during diastole Its physiological bases are *1*) that the oxygen requirement of the heart is decreased by reduction of afterload and 2) that coronary volume flow can be greatly increased by a high diastolic perfusion pressure which acts when ventricular wall tension is low and, consequently, coronary vascular compression is minimal *Synchronized arterial assistance* (Fig 3 134 *C*) consists of withdrawal of blood from the aorta just before and during systole, thus reducing the pressure against which the left ventricle ejects blood, and returning the same volume of blood during diastole, to increase diastolic pressure and thereby increase coronary flow

Intra-aortic counterpulsation (Fig 3 134 *D*) is a blood volume displacement method A gas-filled balloon in the aorta is deflated during cardiac systole, thus lowering intra-aortic pressure during left ventricular ejection The balloon is then distended by gas pressure after aortic valve closure, this increases diastolic pressure and coronary flow by volume displacement

Aortic blood volume displacement is also involved in *external limb counterpulsation* (Fig 3 134 *E*) The limbs are incased in a sheath so that when suction is applied during systole, arterial capacity in the limbs is increased, with drawing blood from the aorta Positive pressure during diastole returns blood to the aorta

Right ventricular assist by synchronous respiration (Fig 3 134 *F*) With this method the capacity of the pulmonary vascular bed is increased during cardiac systole so that right ventricular afterload is decreased This is accomplished by pressure variations applied either through an airway or externally to the chest wall by a cuirass The negative pressure is synchronized with systole, increasing pulmonary vascular bed capacity and decreasing right ventricular pressure

Ballistocardiogram-triggered assistance (Fig 3 134 *G*) has been tried with the idea of moving the body appropriately during systole and diastole as a counterpulsation method

An "in series" auxiliary ventricle has been employed to accomplish counterpulsation (Fig 3 134 *H*) A pneumatically activated pumping chamber, which contains no valves, is inserted in series with the aorta During systole the pumping chamber is evacuated to lower aortic pressure (afterload) During diastole, after the aortic valves close, the blood is pumped back into the aorta, increasing flow into the coronary arteries and the rest of the circulation

An *"in-parallel" auxiliary ventricle* (Fig 3 134 *I*) has been developed which takes blood from the left atrium and returns it into the aorta during diastole Thus left ventricular work is decreased and coronary flow increased This is a form of left ventricular bypass Another type of left ventricular bypass via the peripheral circulation similarly accomplishes "parallel pumping" (Fig 3 134 *J*) Blood is withdrawn from the left atrium into a reservoir and pumped back into the arterial system through a femoral artery If the blood is withdrawn rapidly enough, left ventricular filling may be prevented and full bypass of the left ventricle accomplished

Right ventricular bypass (Fig 3 134 *K*) is effected by removal of blood from the right atrium and pumping it back into the pulmonary artery, thus sparing the right ventricle In *complete heart-lung bypass* (Fig 3 134 *L*), the systemic venous return is drained into a reservoir, pumped through an oxygenator and returned to the arterial system under pressure Coronary sinus blood returning to the right

atrium and bronchial flow drainage into the left atrium are removed via catheters

Veno-arterial bypass with oxygenation (Fig 3 134 *M*) is a form of heart lung bypass in which blood is taken from the venous side, oxygenated, and pumped back into the femoral artery, either in an unsynchronized manner or coordinated with the heartbeat (e g , *veno-arterial phased pulsatile partial bypass*, Fig 3 134 *N*)

Finally, implantable devices which can supplant the heart mechanically have been devised and tested in animals

Except for short term use, as in cardiopulmonary bypass for heart surgery, these various methods of circulatory assistance are in the experimental stage Some require formidable surgical intervention, many produce serious hemolysis, and lead to thrombosis and embolization through effects on blood clotting factors Although their practical application lies in the future, there can be no doubt of the enormous benefits successful devices will confer in the treatment of cardiac failure and circulatory shock (see Soroff et al , 1969, Proctor, 1969, Kolff et al , 1969, and Burns et al , 1969)

Various expedients have been used to perfuse an organ or region (Fig 3 134, *O*) For example one end of a pump may be connected to a local arterial source and the other end of the same pump to the artery of the organ or region being supplied Either the perfusion pressure or flow rate can be varied separately This approach obviously has the important advantage that it enables the investigator to study separately the peripheral factors in the organ or region that regulate flow

MEAN FLOW RECORDERS

Because of the imposed resistance to flow, most mean flow recorders are not suitable for use in the path of venous flow but are acceptable in the arterial system where a small drop in pressure across the meter is not so important The types that have given much information regarding the operation of the cardiovascular system include the mechanical stromuhr, bubble flow meter, rotameter, and turbinometer How they work is shown in Fig 3 135*A*, *B*, and *C* They have the disadvantages that they require anesthesia, injection of anticoagulant, insertion between the cut ends of the vessel in which flow is being measured, and their response time is slow, the fastest (the rotameter) being of the order of 90% complete in 1 sec Despite these disadvantages and although arterial flow is highly phasic these devices have been most helpful

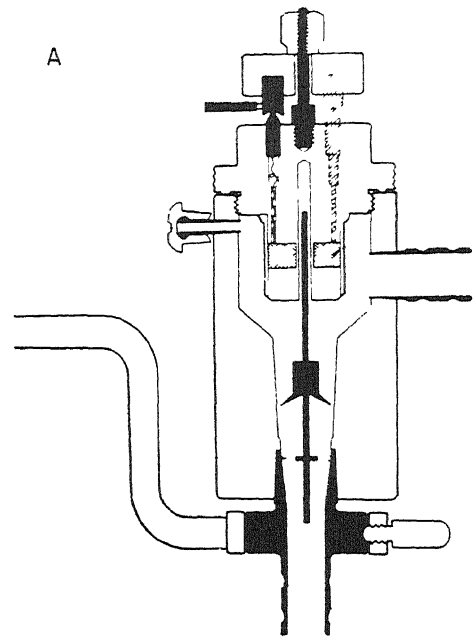
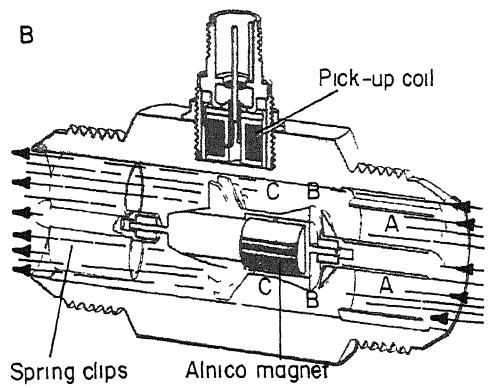


Fig 3 135 *A* sectional view of a rotometer Blood flows through the flow chamber entering the bottom inlet tube and leaving through the outlet tube (*right*) supporting the cylindrical float Float position is detected visually or electrically by an induction coil through which a soft iron upper rod extension of the float projects (Shipley and Wilson 1951)



B sectional view of the Potter turbinometer Flowing blood turns the turbine which contains a magnet The rotating magnet induces an electrical signal in the pick-up coil The amplitude of the induced signal is proportional to the rate of rotation and is calibrated for various flow rates (Sarnoff and Berglund 1953)

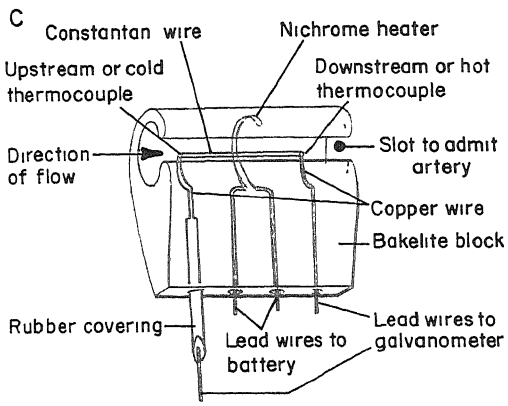


Fig 3 135 *C* diagram of direct current thermostromuhr Blood flowing in a vessel placed in the sleeve reduces the heat created by the electrical heater Temperature changes are detected by the thermocouple unit and reflect rate of blood flow The device is unreliable if external factors alter the relative rate of heat loss (Gregg 1950)

PHASIC FLOW METERS

These devices faithfully record the phasic or moment to moment changes in arterial and/or venous flow during a cardiac cycle (Fig 3 136) The differential pressure and pendulum types must be used with heparin and inserted between the cut ends of a vessel, the electromagnetic and sonar flow meters are used to measure blood flow in intact vessels in unanesthetized active animals, the electromagnetic flow meter has been used in anesthetized humans

Differential Pressure Flow Meters By introducing a device between two points in a flowing stream, a pressure difference is created which varies with the velocity of blood flow The ways of creating the pressure differences are illustrated (Fig 3 136*A*)

Pendulum Flow Meters In a principle first described in the 15th century, a small pendulum, paddle, reed, wire, or bristle which is introduced into a blood vessel is deviated from its resting position by the flowing blood, the amount of deviation being proportional to the velocity of flow

Sonar Flow Meter An electrosonic meter has been used to measure flow in the arteries, aorta, and vena cava, and to measure left ventricular diameter of unanesthetized dogs (Fig 3.136 C) In principle, the transit time of sound moving downstream is determined by the velocity of sound in the motionless medium plus the velocity of the medium The transit time of the sound upstream is determined by the velocity of sound in the motionless medium minus the velocity of the medium Thus, the difference in transit time of the ultrasonic waves upstream and downstream is proportional to the velocity of the medium In practice, barium titrate crystals are mounted at diagonally opposed positions near the end of a lucite cylinder divided longitudinally and this is clamped about an artery Bursts of ultrasonic waves (3 megacycles/sec) emitted by one crystal pass through the vessel wall and diagonally through the stream of flowing blood to reach the opposite crystal Bursts are emitted at about 12,000/sec with direction reversal at 800/sec The difference in transit time of the ultrasonic waves is converted into a direct current voltage, amplified and recorded By turning off the ultrasonic pulse generator, the flow beam reaches a position which is the same as that following temporary mechanical occlusion of the vessel This device appears to be usable on both artery and vein

Electromagnetic Flow Meter This type has been used for pulsatile flow recording in the aorta and arteries of the dog and man (Fig 3 136 D and E) A number of different types have been produced, some commercially In the author's laboratory, the Kolin type, somewhat modified, has been used In principle, an electromotive force or voltage is induced in a conductor moving through a magnetic field at right angles to the lines of force In biological use, blood moving within an intact blood vessel serves as the conductor and the induced voltage which is proportional to the blood velocity is picked up by a pair of electrodes positioned against the vessel wall, perpendicular to the magnetic field and to the direction of blood flow The flow meter which is encased in plastic, consists of a miniature electromagnet made up of coils of wire mounted on steel laminations, a slotted plastic tube into which the vessel is inserted, two electrodes, and wires for energizing the electromagnet and for leading off the induced voltage (one-millionth volt) which is amplified and recorded In use, the probe is implanted around an intact blood vessel by temporarily decompressing and forcing the vessel through the longitudinal slit in the probe Its main advantages are its ability to

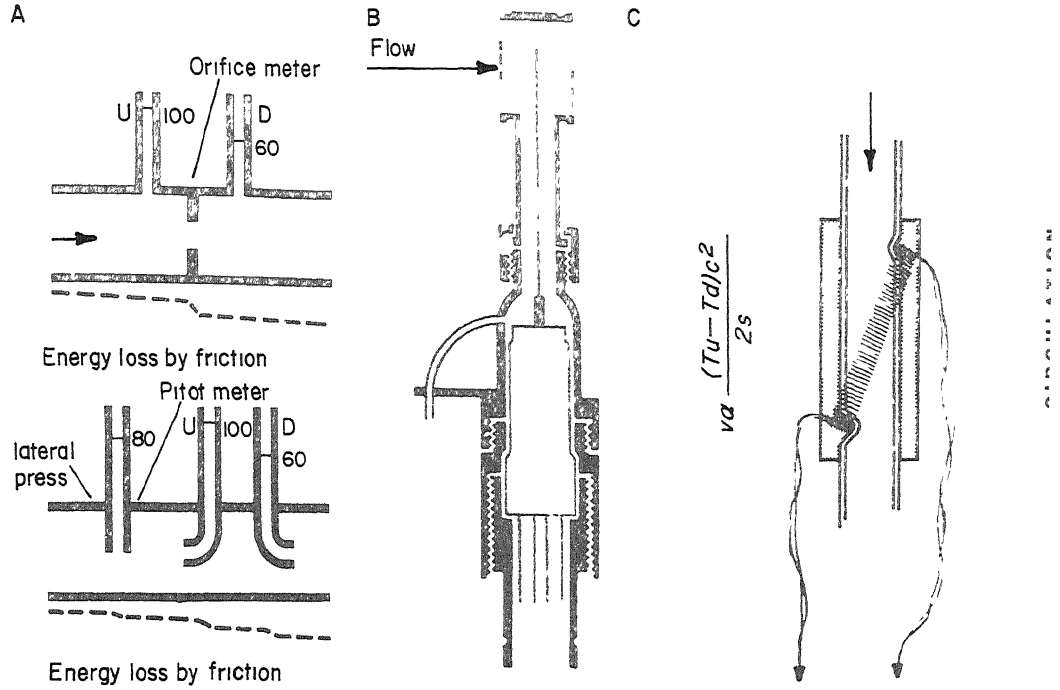
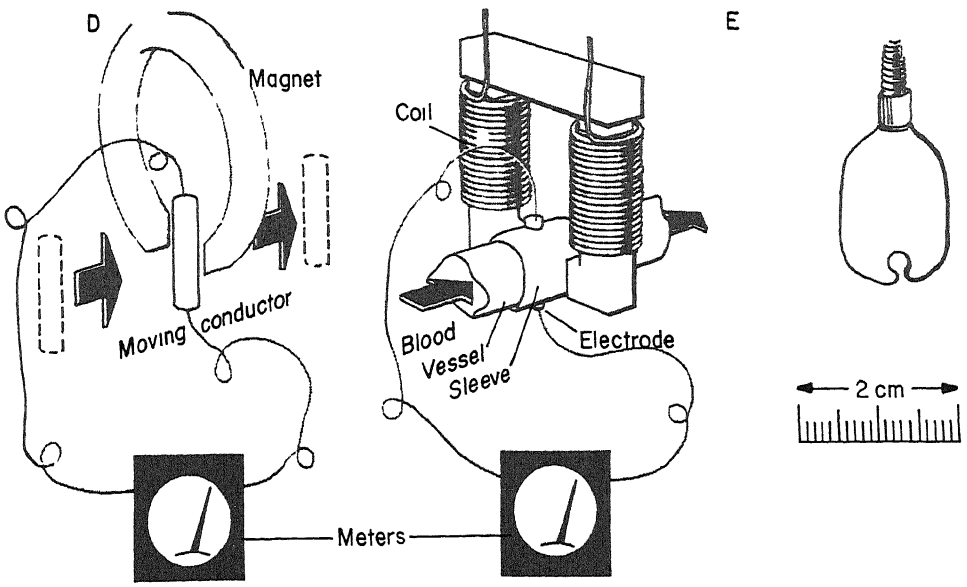


Fig 3 136 ***A B C*** Phasic blood flow meters ***A,*** the orifice meter and the Pitot tube In the orifice flow meter the rate of flow is proportional to the difference in pressure proximal and distal to the region of constriction In the Pitot tube pressure differences are detected between a tube facing upstream and one facing downstream (Brecher 1956) ***B,*** the bristle flow meter Flowing blood deviates a small bristle The degree of deviation is detected electronically and serves as a measure of blood flow (Brecher 1956) ***C*** sonar flow meter Ultrasonic waves are beamed between two crystals Velocity of sound transmission is proportional to the velocity of the blood (Franklin et al 1959) ***D,*** electro magnetic flow meter principle ***E,*** electromagnetic flow meter probe



measure mean and phasic flow in unopened large and small vessels of chronic animals, its linearity of calibration and its high frequency of response The flow beam can be made to approximate the zero flow position by "electronic zeroing" or repeated mechanical flow zeros can be easily obtained even in the exercising dog by temporary (3 to 4 sec) inflation of a rubber pneumatic cuff just distal to the flow meter

FICK PRINCIPLE

Much of our information concerning blood flow, especially the output of the heart is derived from the use of methods based on the Fick principle To employ the Fick principle to determine blood flow through the whole body or an organ, it is necessary to have a substance that is either removed from or added to the blood during its period of flow through the vessels under study The amount which is added or removed from the organ by the blood is equal to the difference between the amount brought into the organ and the amount carried away from the organ The amount of the substance in the blood is expressed as the volume multiplied by the concentration Thus $Q = F_1 \times C_1 - F_0 \times C_0$ in which Q = amount of reference substance picked up or released by the organ, F_1 = blood flow into organ, F_0 = blood flow out of organ, C_1 = concentration of reference substance in the inflowing blood, and C_0 = concentration of reference substance in outflow blood Since the inflowing and outflowing volumes are assumed to be the same, the equation becomes

$$F = Q/(C_0 - C_1) = Q/\Delta C$$

For Cardiac Output with Oxygen and Carbon Dioxide. Measurement of the output of blood by the animal or human heart has been calculated from the difference between the oxygen content of the mixed venous blood and that of the arterial blood and the total oxygen consumption of the body It is evident that if the quantity of oxygen which a unit of blood delivers to the tissues (or takes up from the lungs) is known, together with the total quantity of oxygen consumed over a given period, then the volume of blood which had been engaged in the carriage of this quantity of gas can be calculated For example, the arterial blood contains about 19 volumes of oxygen per 100 ml. It gives up, let us say, 6 volumes to the tissues, i e , the mixed venous blood coming to the lungs contains 13 volumes % The arteriovenous (AV) oxygen difference is, therefore, 6 volumes % If the total quantity of oxygen consumed per minute is found to be 250 ml , then the cardiac output is

$$[250 \text{ ml/min} / (19 - 13 \text{ ml}/100 \text{ ml})] \times 100 = 4\ 2 \text{ liters/min}$$

The oxygen consumption must be measured very accurately for several minutes and the subject must be in a steady state The blood samples for determining arteriovenous oxygen difference should be obtained simultaneously Arterial blood may be obtained from any artery but only *mixed* venous blood can be used since the quantity of oxygen contained in venous blood varies grossly depending upon the vascular bed from which it drains Even the oxygen content of blood in the superior vena cava differs from that in the inferior vena cava by 1 to 2 volumes % Mixing of the blood occurs in the right ventricle and is generally complete by the time the blood reaches the pulmonary artery where sampling must be made for accurate determinations

The cardiac output can be calculated in a similar way from the total carbon dioxide elimination and the arteriovenous carbon dioxide difference (CO_2 in mixed venous blood less CO_2 in arterial blood) However, the vagaries of cardiac output calculation on this basis are such that the measurement is generally ignored The unreliability of the Fick method when carbon dioxide is used is related to the fact that small changes in ventilation strongly influence carbon dioxide storage by the body and a steady state is hard to reach

The direct Fick method was first employed in man by Klein (1930), following Forssman's (1929) demonstration upon himself of the safety of right heart catheterization Cournand and his associates (1941) established the safety of the procedure and stimulated widespread utilization of the method The technique of cardiac catheterization has been described in detail by Cournand and others A sample of mixed venous blood is drawn from the pulmonary artery by means of a catheter passed up an appropriate arm vein The introduction of the catheter is guided by x-ray visualization (fluoroscope) A sample of arterial blood is obtained by arterial puncture The blood samples can be analyzed for their oxygen contents, and after the total oxygen consumption has been deter-

mined in the usual way, calculation of the cardiac output is made from the formula already given The oxygen contents can be measured with the Van Slyke apparatus or, for more rapid determination, a photoelectric method has been developed (Fig 3 137) Blood for analysis is drawn through a cuvette which is transilluminated by a constant intensity light source The light passing through a 628 filter falls upon a photomultiplier tube

The method presupposes that the oxygen consumption and the AV oxygen difference are obtained simultaneously and are constant during the time of measurement Both cardiac and respiratory cyclic changes in the concentration difference of oxygen and the flow rate are capable of introducing large errors in cardiac output calculations These changes do not appear to occur to any significant degree in man when a steady state has been achieved A potentially serious error of the Fick application may occur if gas is stored in, or liberated from, the body, including the lungs This would occur when the relation of aerobic to anaerobic metabolism of the body is altered and, as a consequence, the level of blood oxygen is altered without relation to blood flow Thus, the Fick calculation is dubiously related to cardiac output as the organism goes into or comes out of an anoxic state or in the presence of congenital heart disease with cyanotic episodes In the latter instance, the shunts between the pulmonary and systemic circuits induce large errors in calculation of systemic flow Special formulas have been developed to calculate the volume of blood flow through such shunts

Cardiac output determined by the Fick method probably has an error of $\pm 10\%$ when very carefully done

For Cardiac Output with Foreign Gases. Most of the early figures for cardiac output were based on the Fick principle using a foreign gas For this, the lungs are used as a gas aerotonometer to measure the gas tensions of mixed venous blood and hence, its gas content A subject breathes an inert foreign gas (i e, one which becomes dissolved in plasma but does not combine with hemoglobin, lipoid or other constituents of the blood) If the rate of gas absorption by the blood, its alveolar tension and its stability in blood are known, then the quantity of blood which has passed through the lungs can be calculated

$$\text{Flow} = \frac{\text{rate of gas absorption}}{\text{alveolar gas conc} \times \alpha_B}$$

where α_B is the solubility coefficient of the gas in question in blood

This is the principle on which a foreign gas method is based Nitrous oxide (N_2O), ethyl iodide (C_2H_5I), and especially acetylene (C_2H_2) have been used in the past The main difficulty of the method is that equilibrium between alveolar air and the venous blood must be effected before recirculation Originally, it was felt that the circulation time was 25 to 30 sec, and hence, that a respiratory mixture could be left in the lungs for that length of time without

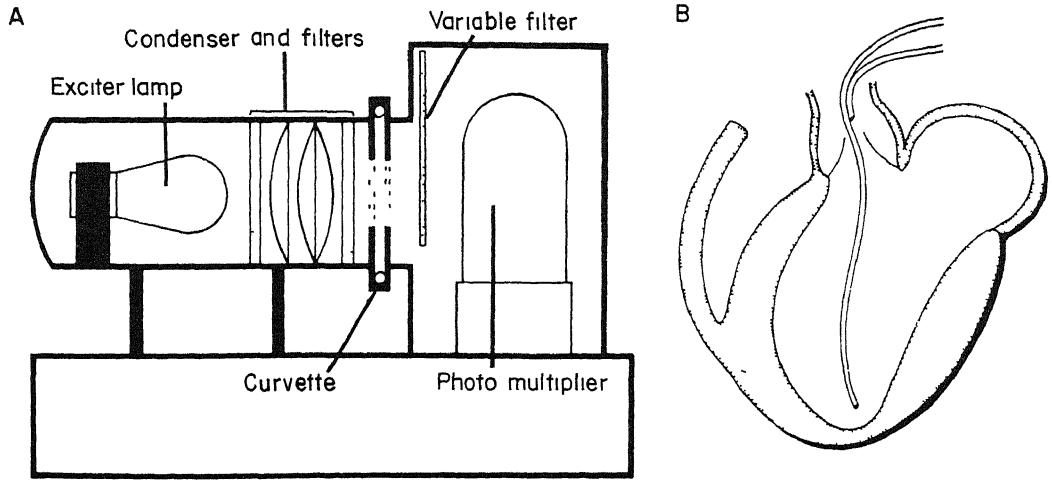


Fig 3 137 *Left* drawing showing principle of the densitometer used to determine instantaneous cardiac output and continuous blood O_2 saturation *Right* principle by which change in blood conductivity in a ventricular cavity is used to determine diastolic ventricular volume and stroke volume

exposure to recirculated blood However, the correct figure for circulation time is much shorter, 10 to 18 sec , which is often an insufficient time for equilibrium to be established, hence, the values for resting cardiac output established by this method are usually lower than normal and, at times, grossly in error Nitrous oxide has been used to estimate not only cardiac output but the instantaneous capillary blood flow since it is so soluble that it is immediately picked up by blood entering the capillary bed of the lung (Fig 3 138) If a subject is in an airtight chamber and takes and holds a single breath of N_2O for 8 to 10 sec , the chamber pressure falls at a rate which indicates the rate of absorption of the gas by the blood and, consequently, the rate of pulmonary capillary blood flow or cardiac output Recirculation is not a problem during this short time The instantaneous rate of blood flow can be calculated since the rate of N_2O absorption, the alveolar gas tension and its blood solubility are known

For Regional Blood Flow. Of the methods considered, except for the electromagnetic flow meter and sonar techniques, measurements of volume of blood flow passing through the splanchnic bed of kidney cannot be made directly even in experimental animals without destructive surgical procedures or unphysiological interference with function The Fick principle has been used to determine mean blood flow in man and animal through the liver, and kidneys, as well as the coronary and cerebral circulation The introduction of venous catheterization has made it possible to obtain blood issuing from the hepatic vein in intact animals and man, thus permitting the development of methods of measuring flow on the basis of the Fick principle Bromsulphalein (BSP) has proved especially useful here because it is taken up by the liver with remarkable efficiency Of the total removed from the blood each minute, when the plasma BSP concentration is maintained by constant intravenous infusion below 2 or 3 mg /100 ml , no more than 5% can be attributed to escape through extrasplanchnic portals For practical purposes, therefore, the hepatic removal of BSP equals the rate of infusion as long as the plasma level is constant The splanchnic blood flow may be calculated by dividing the removal rate by the difference between BSP concentrations in peripheral arterial and hepatic venous blood The flow value is

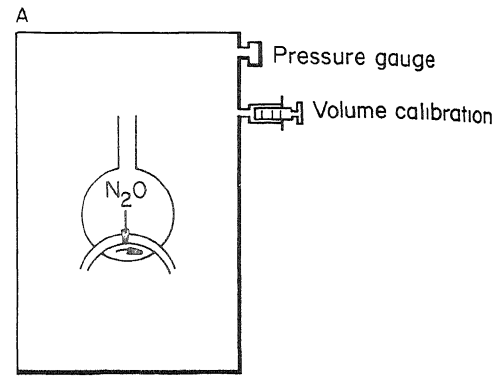
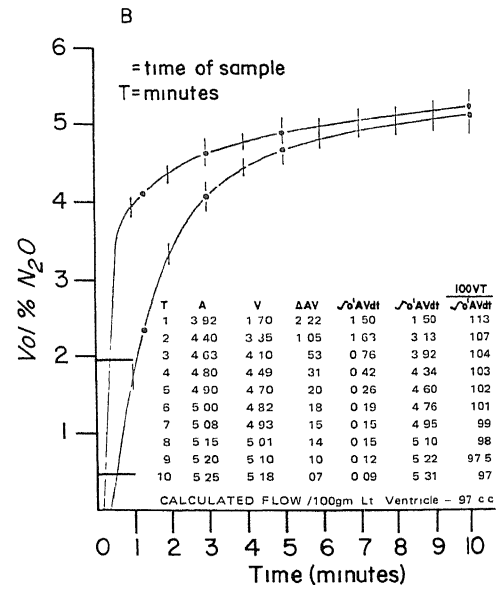


Fig 3 138 *A* diagram illustrating the principle of measurement of pulmonary capillary blood flow (A B DuBois and R Marshall J Clin Invest 36 1957 1566) *B* graph indicating typical nitrous oxide saturation curves obtained in determining coronary flow per 100 g per left ventricle per min in dog *Upper curve* from carotid artery *lower curve* from coronary sinus by catheter (Gregg 1950)



T	A	V	ΔAV	√ΔAVdt	√ΔAVdt	100VT / √ΔAVdt
1	3 92	1 70	2 22	1 50	1 50	113
2	4 40	3 35	1 05	1 03	3 13	107
3	4 63	4 10	53	0 76	3 92	104
4	4 80	4 49	31	0 42	4 34	103
5	4 90	4 70	20	0 26	4 60	102
6	5 00	4 82	18	0 19	4 76	101
7	5 08	4 93	15	0 15	4 95	99
8	5 15	5 01	14	0 15	5 10	98
9	5 20	5 10	10	0 12	5 22	97 5
10	5 25	5 18	07	0 09	5 31	97

generally referred to as the estimated hepatic blood flow since sampling is made in only one of several hepatic veins A number of other substances such as Rose Bengal and ^{131}I, have been used in the same way to measure splanchnic blood volume

Similarly, the blood flow through the human kidney can be determined Diodrast, when in low concentration in the plasma, undergoes virtually complete clearance, i e , it is removed from the plasma in a single passage of blood through the kidney This substance is excreted

largely by the tubules, only about 16% is filtered through the glomeruli When the amount excreted in the urine within a time and the plasma concentration are known, the quantity of blood that has passed through active renal tissue can be calculated For example, if 600 mg of Diodrast are excreted in the urine per minute, and each milliliter of plasma contains 1 mg, then to deliver 600 mg of Diodrast, 600 ml of plasma went through the kidney, or with a hematocrit of 40, about 1000 ml of blood flow

Mean blood flow has been successfully determined through the coronary circulation with nitrous oxide inhalation and through the cerebral circulation with nitrous oxide or krypton inhalation It is the only method available for such studies in man (Fig 3 138) In use in the heart or brain, several simultaneous arterial (any artery) and venous (coronary sinus for the heart and the internal jugular vein for the brain) samples are drawn from the beginning of N_2O inhalation (mixture of nitrogen, 64%, N_2O, 15%, O_2, 21%) to the time of its equilibrium in the bloodstream or during denitrogenation after previous inhalation of the nitrogen mixture Continuous simultaneous samples can also be drawn from the artery and vein for analysis The flow per unit of time through these organs is equal to the amount of the substance taken up by that organ in a given time divided by the difference in concentration of the substance in the arterial blood and venous drainage of the organ in the same time period The denominator in the Fick equation is found by computing the integrated difference between the concentrations of N_2O in arterial and venous blood during the period of equilibration with a low concentration of N_2O The concentration of gas in the tissue at the time of equilibration (the numerator in the Fick equation) is unobtainable directly and is assumed to be equal to the product of the venous blood concentration of the gas (after equilibrium is established) and a partition coefficient (1 0 in case of the heart and 1 3 for the brain) When the equation is multiplied by 100, units for blood flow are obtained which are expressed as milliliters of flow per 100 g of heart or myocardium

INDICATOR DILUTION PRINCIPLE

The principle is based on observation of the rate at which a known quantity of material introduced into a vessel or heart cavity is diluted by blood flow through the vessel or cavity In the principle as originally applied by Stewart and modified by Hamilton, it has been shown that the average rate of blood flow can be stated

$$F = E/\int_0^t C\,dt$$

where F = average rate of flow, E = total quantity injected, C = instantaneous concentration of the injected material From this relationship, it is possible to determine the average flow rate in almost any cavity or vessel of reasonable size if the amount of injected material is known and if the concentration-time function can be accurately determined (Fig 3 139)

Blood Velocity

Of the two measurements, volume flow of blood in milliliters per minute and the velocity

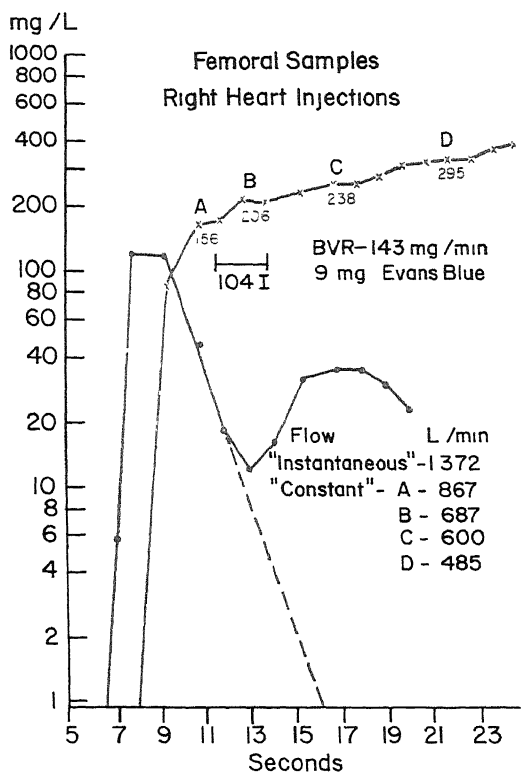


Fig 3 139 Concentration curves resulting from continuous infusions of brilliant vital red (BVR) and rapid injection of T1824 (Evans blue) both started at zero seconds Short horizontal line and figure denote the expected concentration plateau level of the continuous infusion on the basis of instantaneous injection (I) (Howard et al 1953)

of flow in centimeters per second, the former is probably more important The major function of the circulation is to supply the metabolic needs of the tissues and volume flow is closely related to this, whereas measurement of velocity of flow alone does not necessarily tell us anything about the blood supply to the tissues The relation between the observed velocity and volume flow is not consistent For example, in the presence of local constriction in a vessel without constriction in other areas, flow velocity will increase, but flow will decrease In the presence of general as well as local constriction, velocity and volume flow will both decrease Despite this, knowledge of the flow velocity is highly important to an understanding of hemodynamics of the circulation Flow velocity measurements have been made under direct vision, by use of the various flow meters that have been described, by observation of movement of injected substances, and finally, by determination of the differential pressure within an unopened and unrestrained vessel

Visualization. There is a natural tendency to assume that when one actually sees the movement of blood within a vessel, this must supply the most certain evidence of blood velocity However, when this is done, one tends to be impressed by the movement of the red cells and not by that of the blood as a whole

Flow Meters. By insertion into or application onto a blood vessel of almost any of the flow meters described, one automatically measures the velocity of flow since the meter insures a fixed diameter through which the flow occurs

Movement of Injected Materials. Various noteworthy attempts have been made to follow cyclic changes in velocity in the unopened vessel by recording the movements of various injected foreign materials such as air bubbles, dyes, and x-ray opaque fluid The technique is difficult since the moving front of the injected material, i e , the difference of optical density between blood and the contrasting medium, is often difficult to detect This is because the velocity is not the same at different points along the radius of the blood column, the flow being much greater in the axial part of the stream than toward the periphery, and also because the flow is not always unidirectional A mean or average velocity is obtained which is derived from the flow indicated by mean flow meters. It, of course, ignores very large fluctuations in velocity which occur during a cardiac cycle The moment-to-moment or phasic velocity is recorded by those pulsatile flow meters with a frequency of response sufficiently rapid to follow it

Differential Pressure Velocity The instantaneous blood velocity has been determined in the aorta and arteries from the instantaneous pressure difference between two points located along the axis of flow or along the vessel (Sect 3, chap 1, Fig 3 29) Such pressure gradients have been measured by simultaneous recording of arterial pressure at the two points either with individual pressure manometers or by use of a double lumen catheter attached to a differential pressure strain gauge In the double lumen catheter which has been used in the aorta of humans, the lateral opening of one catheter tip opens at a point about 5 cm from the other From the pressure gradients, velocity curves and flow curves have been calculated which agree well with those directly measured

A rough approximation to mean linear aortic velocity in man can be made using the cardiac output and the cross-sectional area of the aorta (measured in a cadaver of the same size)

The mean velocity ($\bar{v}$), in cm /sec equals cardiac output (V) in ml /sec /cross-sectional area (A) in cm^2

$$\bar{v} = \dot{V}/A$$

For a human aorta of radius 1 0 cm and cardiac output 5 liters/min , the mean velocity would be

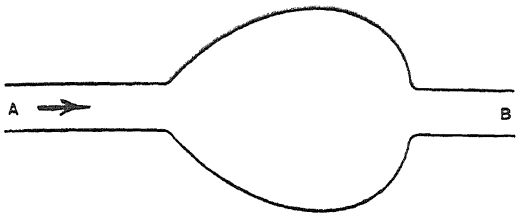
$$\bar{v}\,\text{aorta} = 83/3 = 28\ \text{cm /sec}$$

This is *mean* flow Since flow is pulsatile, the velocity at systolic peak is perhaps three times this value and much less at diastolic nadir

CIRCULATION TIME (C T)

Circulation time is defined as the shortest time which a particle of blood takes to go from one point in the circulation to another Circulation time is most often determined for a portion of the circulation, such as from the right to left side of the heart

The circulation time of a particle of a substance can be accurately determined if it is injected instantaneously into a blood vessel, if it moves at the same rate as the blood stream and if its arrival further down the vessel is accurately detected Practically, the situation is much more complicated. When a reasonable amount of a substance is injected, its arrival



$$\text{CIRC TIME IN SEC} = \frac{\text{VOL A TO B}}{\text{FLOW IN ml/SEC}}$$

Fig 3 140 Schematic diagram illustrating the relation between circulation time volume flow and the volume of the vascular bed

time is so spread out at the downstream detection point that detection is difficult If the circulation time is made across an organ or a major portion of the circulatory system, the spreadout is even greater In addition, if all the blood channels between the two points are not approximately uniform in length and diameter, only a part of the blood will pass from point to point at the rate indicated by the measurement Despite these considerations, measurements of circulation time can be semiquantitative

The clinical methods for estimating circulation time use injected substances, such as Decholin and ether, that signal their arrival by effects on special receptors, such as the taste budds in the tongue With Decholin, the normal range of arm to tongue (bitter taste) time is 8 to 17 sec , with ether, the normal arm to lung (patient first smells ether) is from 4 to 8 sec For greater precision in experimental investigations, these classical methods have been superseded by the injection of dyes and radioactive tracers which can be easily and objectively detected by densitometers, conductivity cells and isotope counters The main technical difficulty is the proper calculation of the average transit time In a symmetrical curve, the mean (average of all transit times), the mode (time coordinate of the peak), and the median (time coordinate which halves the area) all coincide But as the curve becomes asymmetrical, and particularly as the terminal portion stretches out more in time than the initial part, the three measures depart more and more from each other The mean circulation time (MCT) it takes all dye to pass is computed as the center of gravity of the curve

The significance of a change in circulation time is debatable The circulation time between two points in the vascular system is equal to the volume of the vessels between the two points divided by the flow (Fig 3 140) The validity of this relationship has been checked in glass models and mathematically It is obvious that a change in circulation time could be due to a change in volume of the vascular bed between the point of injection and sampling, as well as to change in flow Since change in flow follows any vasomotor change, and since the resultant redistribution of pressures causes local volume change, both numerator and denominator of the ratio for circulation time (volume over flow) change simultaneously They generally change in the same direction, and so the circulation time could be a very poor indicator of blood flow, volume of blood, or vasomotor change

3 / CONTROL MECHANISMS OF THE CIRCULATORY SYSTEM

Functional Organization of the Vascular Bed

The sequential segments of the circulatory system are functionally differentiated as illustrated in Figure 3 141 These series-coupled segments make up the vascular beds for the various tissues The circulatory system consists of a set of these circuits coupled in parallel (Fig 3 142)

Vessels may be classified in accordance with their function as follows (Folkow, 1960)

1) Windkessel vessels, which convert pulsatile inflow to a somewhat smoothed outflow
2) Resistance vessels, chiefly the small arteries, the arterioles, and to a lesser extent the capillaries and smallest vessels
   a) Precapillary resistance resides chiefly in the small arteries and arterioles, it is the major site of resistance to flow
   b) Postcapillary resistance is determined by the venules and veins Capillary hydrostatic pressure and, thus, filtration-absorption exchange, is determined by the ratio of precapillary to postcapillary resistance
3) Sphincter vessels, terminal segments of precapillary arterioles which can constrict and restrict or shut off capillary flow They determine the number of open capillaries and hence the area of capillary exchange surface
4) Exchange vessels, the capillaries, which are not contractile, but respond passively to changes in resistance and sphincter vessels
5) Capacitance vessels, chiefly veins which through diameter changes having little influence on resistance can effect marked shifts of blood volume and thus affect cardiac venous return and volume flow dramatically They contain about 80% of regional blood volume
6) Shunt vessels arteriovenous anastomoses found in some tissues, where they allow bypass of exchange vessels

Resistance and capacitance vessels may react differently in accordance with their differing functions Thus, epinephrine dilates some resistance vessels while constricting capacity vessels, norepinephrine constricts both, but affects capacity vessels most

Control of Vascular Smooth Muscle

LOCAL CONTROL

The smallest blood vessels have an inherent, myogenic automaticity or basal degree of contraction of their contained smooth muscle which is independent of blood borne substances or nervous influences, but which can be altered by stimulation of nerve fibers, local electrical stimulation, or by changing the physical and chemical environment The mechanism of this inherent vasomotion is unknown Presumably, it arises from locally produced metabolites which can have complex effects Locally, about the arterioles, precapillaries and capillaries, there are released norepinephrine, acetylcholine, CO_2, lactic acid, products of nuclear metabolism, and presumably many other substances, which can act in many different ways on these vessels

Several potent vasoactive substances such as histamine, prostaglandins, serotonin, and adenosine tri- and diphosphate are found in various tissues, but their importance in local vascular regulation has not been established The extent of such basal vasomotion has been evaluated in sympathectomized, vascular areas where neurohumoral, constrictor influences are eliminated

Although large quantitative differences exist in different parts of the vascular tree, generally speaking there is an inverse relation between the local, inherent activity of smooth muscle cells and the extent of neurohumoral control in any given vascular bed The cutaneous arteriovenous anastomoses, such as in the paw of the

cat, which are strongly controlled from the hypothalamic heat loss center, are maximally dilated when neurohumoral agents are eliminated, while elimination of the vasomotor nerves serving vitally important metabolic areas, such as the brain and myocardium, has scarcely any effect on the local blood flow In areas such as skeletal muscle, neural elimination still leaves a very strong vessel tone in the dog and man, which is increased with time Indeed, some investigators have expressed the belief that in skeletal muscle, blood flow is mainly under local control by metabolites, and not nerves, since subjects with sympathectomized legs do as well running (athletes and normals) before and after sympathectomy Evidence suggests that reactivity to local metabolic effects is stronger in the metarterioles and precapillary sphincters, whereas constrictor fibers predominantly affect the arterioles, indicating that a difference in type of control may exist within the same vascular bed Presumably, in all beds, there is some locally produced vasodilation which counteracts the centrally induced reduction in blood flow by constrictor nerves and protects the tissue against ischemia (see "Autoregulation" p 3-6 et seq)

Myogenic smooth muscle reactions to stretch also operate in local vascular control This has been termed the "*Bayliss mechanism*," since this investigator first proposed that increased stretch might increase vascular tone A variety of experimental observations in vivo and in vitro have confirmed that passive stretch stimulates contractile activity of vascular smooth muscle Two major objections to a regulatory myogenic mechanism have been offered on a priori grounds *1*) Increased vasoconstrictor response to increased wall tension constitutes a positive feedback response leading to instability in the cardiovascular system Thus, increased blood pressure would elevate flow resistance which would further raise the blood pressure, leading to additional vasoconstriction, and so on *2*) Smooth muscle contractile response to vascular distention would tend to abolish the stimulus of stretch The vessel might thus oscillate around a diameter somewhat greater than that at the time of the stretch stimulus These objections are not necessarily valid The possible instability caused by positive feedback could be limited by "metabolic vasodilators" and the maximum range of vasomotor response Considering the structural complexity of passive elastic and viscous elements in vessel walls, the circular, spiral, and longitudinal arrangement of the smooth muscle fibers and the tortuosity of some arteries, it is unlikely that myogenic response to stretch would operate only within rigid diameter limits Thus, both

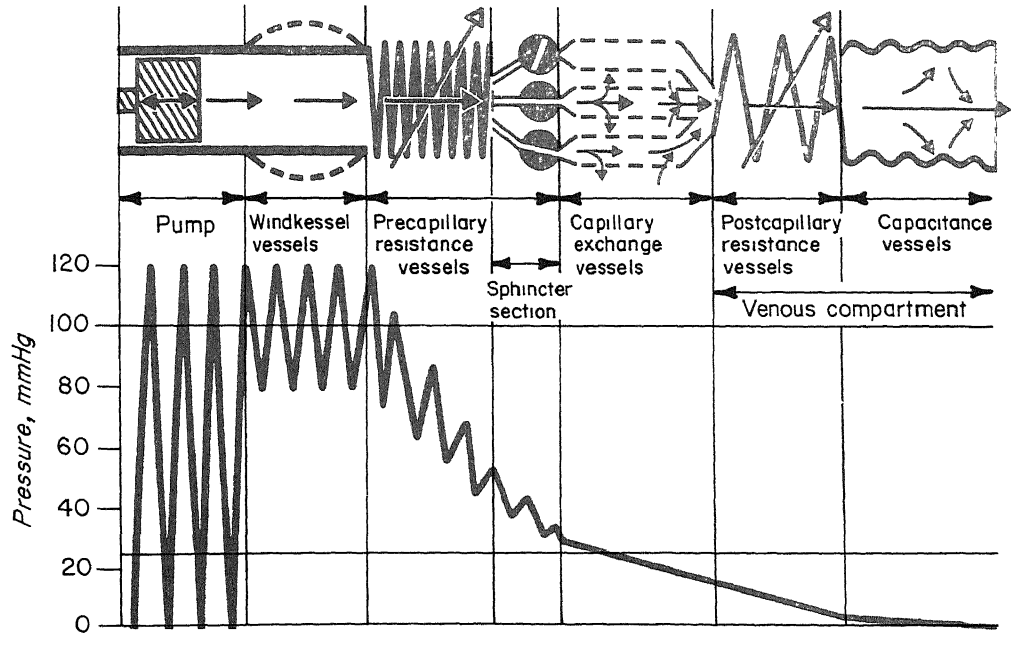


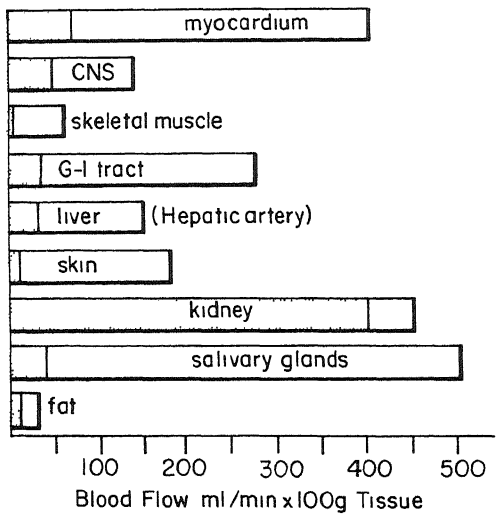
Fig 3 141 Functionally differentiated segments of the vascular bed (Redrawn from Mellander and Johansson 1968)

local metabolic dilator factors and myogenic mechanisms may be envisioned as functioning together to produce autoregulation (Mellander and Johansson, 1968, Mellander, 1970)

SYMPATHETIC VASOCONSTRICTOR FIBERS

Distribution. These were discovered in 1852 by Claude Bernard, who stimulated the cervical sympathetic nerve in the rabbit and observed constriction of the vessels of the ear They belong to the thoracicolumbar (sympathetic) division of the involuntary nervous system The constrictor fibers arise from groups of nerve cells situated in the lateral horns of the spinal gray matter, extending in man from the first thoracic to the second or third lumbar segment, inclusive All the arterioles of the body wher ever situated are supplied with filaments which have their ultimate source in this relatively limited region of the central nervous system They are distributed to the periphery in the manner elsewhere described for the thoracicolumbar outflow in general (see also sect 9, chap 3)

The vascular nerves of the limbs, as shown by Kramer and Todd (1914), are distributed by two distinct modes *1)* A *proximal* innervation arises in the case of the vessels of the upper limb directly from the cervical part of the sympathetic chain—middle and inferior cervical ganglia The fibers pass to the subclavian artery and are conveyed in a plexiform manner along the outer coat of this vessel and its branches, and into the arm along the axillary artery The corresponding supply to the vessels of the lower limb is derived by extension from the aortic plexus in the abdomen The fibers follow the common and external iliac arteries into the thigh The sympathetic fibers derived in the manner just described do not extend beyond the larger vessels of the limbs (proximal portions of the brachial and femoral arteries) *2)* A *distal* innervation is carried to the peripheral vessels via the spinal nerve trunks These nerves reach the arteries at different levels and, penetrating the vascular wall, form a nerve net surrounding the muscular coat, the highest level of this type of innervation probably overlaps the region innervated by the proximal group of fibers mentioned above Adrenergic nerve endings have been identified in all the sequential sections of the vascular beds except the true capillaries Precapillary resistance vessels (small arteries and arterioles) have, in general, a rich innervation In the smallest precapillary vessels the number of fibers is small However, there is experimental evidence for neurogenic constriction of precapillary sphincters The venules have fewer adrenergic fibers than the larger veins, which, themselves, are less richly innervated than the precapillary vessels It is solely through such sympathetic fibers traveling with somatic nerve trunks that constrictor impulses are conveyed to the minute vessels of the limbs Ganglion cells are absent from the vessels of the limbs Section of a peripheral nerve, therefore, causes complete degeneration of vasoconstrictor fibers in the area of its distribution



	Rest blood flow (l/min)	Max blood flow (l/min)	Organ weight (kg)
Myocardium	0 21	1 2	0 3
CNS	0 75	2 1	1 5
Skeletal muscle	0 75	18 0	30 0
G-I tract	0 7	5 5	2 0
Liver (hep artery)	0 5	3 0	1 7
Skin	0 2	3 8	2 1
Kidney	1 2	1 4	0 3
Salivary glands	0 02	0 25	0 05
Fat	0 8	3 0	10 0
Total ≈	5 1	38 0	48 0

Fig 3 142 *Upper* bar graph representing approximate blood flows in various organs at maximal vasodilatation (total bar) and at rest (hatched areas) in ml/min × 100 g tissue at perfusion pressure of 100 mm Hg *Lower,* approximate values for regional blood flows in a 70-kg man at rest' and at maximal dilatation, deduced from organ weights The organs included comprise about 70% of total body weight (Redrawn from Mellander and Johansson 1968)

Vasoconstrictor fibers to the head and neck are conveyed from the sympathetic chain through plexuses investing the blood vessels, but also via peripheral nerve trunks (cervical and certain cranial nerves) The vessels of the abdomen and pelvis are supplied with fibers which pass along the vascular walls from plexuses surrounding the aorta and its branches The sympathetic fibers to the heart arise chiefly in five upper dorsal segments of the cord, pass to the stellate ganglia and upper dorsal ganglia as white rami, and proceed to the heart by a complex plexus Although the coronary vasomotor effect of sympathetic stimulation is overshadowed by metabolic, inotropic, and chronotropic effects associated with vasodilatation, stimulation of sympathetic C factors has been shown to cause active vasodilatation directly when other factors are controlled (see p 3-139)

Site and Mode of Action These nerve fibers constitute a group of powerful vasoconstrictor mechanisms The physiological discharge rate of the vasoconstrictor fibers is 1 to 2/sec to maintain normal vessel tone and reaches 10 impulses/sec , with maximal physiological excitation The chemical transmitter released at the smooth muscle cell is norepinephrine, and norepinephrine is released into the blood stream on intense sympathetic stimulation However, such an overflow does not usually occur at physiological discharge rates These constrictor fibers exert control over the resistance vessels (arterioles and adjacent smaller blood vessels) where the main drop in blood pressure from the arterial to the venous system occurs This can be observed by noting the marked increase in blood flow in a vasoconstricted limb immediately following sympathetic blockade The precapillary resistance vessels are controlled by constrictor fibers, but during prolonged constriction accumulation of "vasodilator metabolites" can oppose neurogenic influence, relaxing precapillary sphincters

In addition, these fibers exert strong control over the heart size and the capacity vessels, mainly the veins, and can alter greatly the venous return to the heart and, thus, markedly influence cardiac output To establish this capacitance action, pressure changes occurring in functionally isolated parts of the superficial and deep venous system of dog and man were used to indicate changes in venous constriction, and evidence of a reflex constriction to a variety of stimuli was obtained Isolated venous segments in the intact forearm constrict after reflex sympathetic stimulation by cold, excitement, etc Venoconstriction in the forearm occurs in response to pooling of blood in the leg Similar observations have been made on the capacity vessels of the splanchnic area These effects are abolished by constrictor fiber blockade by hexamethonium Possibly the most clear-cut experiments in dogs on the quantitative importance of the sympathetic system in adjusting total vascular volume, and hence, venous return and cardiac output, are those in which a pump, with output which remained constant under marked variations of peripheral resistance, was substituted for the left ventricle of a dog The pulmonary drainage of blood was carried to a reservoir by means of a plastic tube in the left atrium The blood was pumped from the reservoir through a T tube in the descending thoracic aorta Under equilibrium conditions (reservoir level remaining constant), if norepinephrine (which simulates sympathetic vasoconstriction in physiological dosages) was injected into the dog, there was not only an increase in vascular pressures but also approximately 250 ml (one-fifth of the dog's blood volume) was forced into the pulmonary drainage reservoir feeding the mechanical left ventricle This could come only from the venous system and the right heart When hexamethonium (simulating a functional sympathectomy) was injected instead of norepinephrine, the pressures fell and the dog took up 200 to 300 ml of blood from the reservoir Further work has extended these findings to man In normal man, removal of large amounts of blood in 50-ml increments produces little change in vascular pressure or hemodynamics But, if done following inhibition of sympathetic vasoconstrictor reflexes with hexamethonium, then with each 50 ml of blood withdrawn, there is a perceptible decline in blood pressure After removal of only 350 to 500 ml of blood, the arterial pressure falls, by decrements, to collapse levels During reinfusion, the process is reversed Thus, sympathetic vasoconstrictor inhibition has converted the vasculature into a static system In addition, using the pressure-volume apparatus of Litter and associates (1954), which can be applied to the extremities of man, it is calculated that infused epinephrine and norepinephrine can shift as much as 500 ml of blood from the veins of dependent limbs in a short period of time

These effects of constrictor fibers on venous volume can considerably change venous return by very small changes in vein caliber For example, if the veins contain 60% of blood volume of which 1 to 2% make up the stroke volume of the ventricles, a 1 to 2% shortening of venous vascular smooth muscle would be estimated to double diastolic inflow to the heart, this same amount of constriction of the resistance vessels would only insignificantly increase the peripheral resistance

Excitation of sympathetic vasoconstrictor fibers affects somewhat the aortic smooth muscle cells and possibly other arteries Presumably, the effect is mainly a change in vascular distensibility and only secondarily are mild pressure changes along the arterial tree produced (see also Chap 11) In general, however, the effect is small In vessels from the root of the aorta down to arteries of 1 mm , induced changes in diameter are probably less than 5% of total diameter even with marked vasoconstriction or vasodilation

Functional Significance. Experimental evidence now indicates that these vasoconstrictor fibers are responsible for blood pressure homeostasis (vascular adjustments derived from the baro- and chemoreceptors), and for the regulation of heat loss by the skin blood flow Thus, the most common type of centrally induced vasodilation is caused simply by an inhibition of vascular tone In fulfilling their role as the main neurogenic adjustors of the peripheral circulation, the vasoconstrictor fibers may show a generalized, strictly segmental, or regional function, depending upon the type of stimulus

Central Control. The control of the degree and sites of vasoconstriction depends predominantly on the influence of hormonal action and on the action of the central nervous system

The Spinal Cord The preganglionic sympathetic neurons may, under some circumstances, exhibit spontaneous activity independent of an excitatory drive from afferent fibers or central levels This spontaneous discharge may arise from a changed oxygen and CO_2 tension which is prone to occur in a spinal preparation in which neuronal respiration and blood supply are without doubt considerably altered

Various afferent impulses are able to call forth spinal vasomotor reflexes, the vasoconstrictor neurons form the efferent link Pain or cold stimulation in the skin induces a segmentally arranged constriction of the intestinal vessels in such spinal animals Vasodilation occurs when the skin is moderately heated More intense and widespread effects are seen in humans with chronic transverse lesions of the spinal cord Guttman and Whitteridge (1947) showed the existence of powerful reflexes, in the spinal man, arising from the hollow viscera In patients with high spinal transection, distension of the bladder would raise the blood pressure from normal levels to 300/140 mm Hg This is ascribed to the action of reflexes initiated by the tension in the bladder wall and mediated by the isolated spinal cord and the sympathetic outflow There could also be demonstrated vasoconstriction in the hand and a fall in calf blood flow These phenomena are consistent with a very widespread vasoconstriction as a result of afferent impulses which spread to an unusual extent in the isolated spinal cord

Medulla The constrictor sympathetic vascular effects are controlled primarily by areas in the medulla oblongata situated in the floor of the fourth ventricle Local electrical stimulation has revealed a lateral "pressor area" and a medial "depressor area" causing vasoconstriction and vasodilation, respectively The vasodilation is caused by inhibition of vasoconstrictor tone, specific vasodilator fibers not being involved This area is defined simply as the vasomotor center but does not represent the highest centers of this system The latter are situated in the hypothalamus and in the cerebral cortex In the intact mammal, complicated afferent influences integrated in the cerebral cortex and hypothalamus modify and control the action of the vasomotor center

The state of activity of the medullary vasomotor center depends upon afferent nerve impulses received from various organs and regions of the body, as well as from other nervous centers, respiratory centers, etc , and upon the chemical composition of the blood In turn, the contractility of the heart, the relative peripheral resistance and, hence, the cardiac output, with its distribution to the various organs, are all mainly controlled by the medulla oblongata responding to impulses from all tissues of the body, including the spinal cord and other portions of the brain itself

The vasomotor center exhibits inherent automaticity, since its continuous discharge goes on even after elimination of all incoming nerve influences Section of the brainstem above the

medulla does not affect blood pressure, indicating that upper areas do not dominate vasomotor outflow from the sympathetic nervous system However, such control may be demonstrated by sectioning the cord in the lower cervical region This interrupts the stream of vasoconstrictor impulses passing from the medullary to the spinal centers, the vessels dilate and the blood pressure falls After a time, however, the blood pressure rises again, the spinal centers exhibit their inherent power of autonomous action, and assuming the duties hitherto exercised by the medullary centers, restore the vessels to their previous state of tonic constriction The time required for the vessels to regain their tone after section of the cord varies considerably in different species

The high degree of vasoconstrictor tone which is normally maintained is shown by the fact that section of the splanchnic nerves doubles the flow in vessels of the denervated area (Burton-Opitz, 1903) After the loss of tone resulting from cord section has been restored (see above) it falls again if the splanchnic nerves are sectioned, but after a time a certain degree of tone is regained This resides in the vascular muscle itself—*peripheral tone* Apparently, a long period is required for the development of the intrinsic arteriolar tone

The Hypothalamus In the hypothalamus, electrical stimulation shows the existence of both excitatory and inhibitory neurons which, presumably, have much to do with blood pressure regulation It must be remembered that drastic redistribution of blood flow following extensive changes in the discharge of the constrictor fibers to some areas is often concealed behind slight blood pressure changes and, hence, blood pressure change alone is not a reliable guide Recently, the hypothalamic area has been explored for structures from which a pure inhibition of sympathetic activity could be induced Such a structure, of very restricted dimensions, has been localized in the anterior part of the hypothalamus, only a few millimeters from the relay station of the sympathetic vasodilator fibers (see next section) From this area, topical stimulation induces a generalized, often very marked, inhibition of sympathetic discharge affecting both resistance and capacitance vessels

An immediate and persistent lowering of blood pressure in both normotensive and renal hypertensive dogs occurs following large bilateral coagulations involving the hypothalamic area This lowering of blood pressure is small in magnitude but is as marked as that resulting permanently from bilateral sympathectomy and low cervical cord transection Whether this lowering in blood pressure has a neural or an endocrine basis has not been determined It is not due to the absence of the neurohypophysial antidiuretic principle, because precipitation of a maximal diabetes insipidus (functional neurohypophysectomy) has been accomplished without affecting blood pressure It is believed not to be due to a neighboring adenohypophysial hypofunction, because the lowering of pressure following adenohypophysectomy is delayed rather than immediate as is the case following hypothalamectomy It is perhaps associated with the immediate and marked decrease in basal energy metabolism which supervenes following appropriate hypothalamectomies

In addition, there is a "heat loss center" in the anterior parts of the hypothalamus which controls the discharge to the vasoconstrictor fibers of the cutaneous blood vessels, and, thus, plays an important role in adjusting blood pressure Electrical stimulation or local cooling of this area brings about a rise in blood pressure (vasoconstriction) while direct warming of this region produces a fall in blood pressure (vasodilation) The cutaneous arterioles and precapillary vessels, and, especially, the arteriovenous anastomoses (shunts), are the vessels most sensitively engaged in control of heat loss However, Keller (1960) has marshalled evidence to indicate that vasoconstrictor impulses responsible for the maintained contraction of cutaneous blood vessels *do not take* origin from the hypothalamus as evidenced by the following experiments An asymmetry in the ear vessels and in skin temperature (Horner's syndrome) does not follow unilateral lesions in the hypothalamus or hemisection of the brainstem at a midbrain, pontile, or upper medullary level, nor does a complete transection of the midbrain or pons results in generalized cutaneous vasodilation Ipsilateral vasodilation does occur following hemisection of the lower medulla and this approximates in magnitude and duration that which follows hemisection of the cervical cord

Diencephalic stimulation in the region of the H_2 fields of Forel in the chloralosed dog produces changes in heart rate and blood flow

distribution similar to those which occur in exercise

Cerebral Cortex Quite a number of studies of cortical vasoconstrictor fiber control are available, but relatively little is known about the circulatory adjustments Stimulation of the motor and premotor cerebral cortex results in marked elevation of blood pressure with constriction of the cutaneous, splanchnic, and renal vessels, and, at the same time, a considerable vasodilation in the skeletal muscle It is believed that these higher centers play significant roles in blood pressure response to pain and anxiety

VASODILATOR FIBERS

Dilator impulses emerge from the central nervous system by *1*) the thoracicolumbar outflow, *2*) the cranial outflow of the parasympathetic division reaching the periphery by way of the chorda tympani, glossopharyngeal, and vagus nerves, *3*) the sacral outflow of the pelvic nerve, and *4*) the posterior spinal nerve roots—antidromic impulses

Sympathetic Vasodilator Fibers. The question of dilator fibers in the sympathetic system has been much debated About 15 years ago, it was generally believed that they took an active part in regulation of vascular tone, their activity being governed by the medulla, and that such fibers were distributed to the skeletal muscles, the facial muscles, the buccal mucosa, the coronaries, the intestines, and to certain skin areas, such as the ear These findings were based on direct stimulation of sympathetic fibers and the recording of volume changes by plethysmography Direct measurement of blood flow (volume changes by plethysmography do not always mean change in blood flow), and location of an area of central control has led to considerable clarification The control of the degree of vasodilation of blood vessels is, indeed, supplemented by a system of efferent dilator nerves transmitting from the central nervous system, but this seems to be distributed only to skeletal muscle Electrical stimulation of the sympathetic fibers usually brings about vasoconstriction in the innervated area, since the action of the vasoconstrictor influence usually predominates over that of vasodilation But, if the vasoconstrictor response is first blocked by α-adrenergic blockade or reserpinization, electrical stimulation will produce vasodilation if the sympathetic outflow contains vasodilator fibers The existence of a sympathetic vasodilator innervation to skeletal muscle of the dog and cat has been established by this technique, but appears to be absent in rabbits and certain other mammals The vasodilator nerves may also be activated by topical stimulation in the brain (hypothalamus), with the result that venous outflow from skeletal muscle of the hind leg of the dog may increase 4 to 5 times (as does the local lactic acid production), but the oxygen usage may decrease, suggesting that shunts are being utilized and capillary flow is reduced In addition, such stimulation results in release of epinephrine from the adrenal glands This increased rate of blood flow approximates maximum dilation achieved by other means, such as acetylcholine injection Since blood flow in the human forearm increases during fainting, a potent means of vasodilation must be present in humans Also, since the blood pressure falls, and the blood flow is greater in the normal forearm than in a nerve-blocked forearm during syncope, this might be explained by active vasodilation in the normal forearm excited by sympathetic fibers to the muscles The transmitter is believed to be acetylcholine and the fibers are exclusively cholinergic, i e , they are blocked by small amounts of atropine They take their origin in the motor cortex, synapse in the hypothalamus, pass as a thin bundle through the ventrolateral portion of the medulla oblongata, and down to the lateral horn of the spinal medulla where they are relayed to lower motor neurons This pathway is anatomically and functionally separate from the medullary vasomotor center which mediates vasoconstrictor activity The functional significance of the vasodilator fibers is unknown, but it has been suggested that they are activated in response to cerebral activity, presumably in association with, or in anticipation of skeletal muscle activity The precapillary sphincters and capacitance vessels in muscle seem to lack this cholinergic innervation Conclusive evidences for the existence of these cholinergic vasodilators fibers in man is lacking

Parasympathetic Vasodilator Fibers. The parasympathetic vasodilator fibers run to restricted cranial and sacral areas such as the cerebral vessels, tongue, salivary glands, external genitalia, and to the bladder and rectum These fibers are probably not concerned with baro- and chemoreceptor control of blood vessels, nor are they tonically active It is

generally believed that these fibers are cholinergic, although their vasodilator effect is very resistant to atropine In the salivary glands and skin there is experimental evidence that nerve stimulation causing cholinergic exocrine gland secretion results in the release of an enzyme which acts on plasma protein to produce the vasodilator polypeptide, bradykinin (see p 3–214) While bradykinin has been proposed as the mediator of vasodilation in association with glandular activity, there is evidence that the role of kinins is only of limited or perhaps no importance (Kellermeyer and Graham, 1968) Their distribution to a few areas with highly specialized functions makes it probable that their function is associated with the special tissue in which they are located

Antidromic Vasodilator Impulses. Stricker (1876) years ago reported that stimulation of the peripheral segments of the cut posterior roots of the sacral nerves caused dilation of the vessels of the dog paw This observation was at variance with the Bell-Magendie law which states that the posterior roots convey only centripetal impulses However, in subsequent years, the former view has been fairly well documented Since such fibers do not convey the induced impulses to the central nervous system, but rather the vasodilator impulses are conveyed along the fibers in a direction opposite to that in which ordinary sensory impulses travel, they are called antidromic. It is the consensus that these "*dorsal root vasodilator fibers*" are not activated reflexly from the central nervous system

The effective stimulus leading to vessel dilation may arise in the skin, to be conveyed antidromically to superficial blood vessels. These are called axon reflexes, but they are obviously not true reflexes since no nerve cell is involved The efferent and afferent limbs of the axon reflex are formed by the branching of a single nerve fiber A stimulus applied to one branch sets up an impulse which travels centrally to the point of division where it is reflected down the other branch to an effector organ The most familiar type of axon reflex is that which involves a sensory nerve fiber and through which vasodilation is effected

Dorsal root dilation is of significance only in tissues with a relatively rich distribution of pain fibers, such as the skin and mucous membranes Various substances have been indicated as the transmitter—a histamine-like substance, acetylcholine, and adenosine triphosphate This local axon reflex in the skin is induced by any factor which causes damage to surface tissues—trauma, cooling, heating, frostbite Thus, this local mechanism contributes to local defense and repair in surface tissues by creating an increase of local blood flow in response to the majority of harmful stimuli

ADRENAL MEDULLA VS NEURAL CONTROL OF BLOOD VESSELS

Aside from direct neurogenic connections to the blood vessels, the presence in the blood stream of circulating catecholamines raises the question of their role in controlling the circulation Presumably they appear through some form of hypothalamic stimulation

Vascular Receptors. The concept of α- and β-adrenergic receptors, although an oversimplification, is useful to explain vascular responses to adrenergic agents α-receptors mediate vasoconstriction and β-receptors mediate vasodilatation The β-receptor is believed to be without innervation and limited to blood vessels of skeletal muscle and myocardium Norepinephrine, which has been shown to be the transmitter of sympathetic vasoconstrictor nerves, acts only on the α-receptors Epinephrine has both α and β effects A third receptor for cholinergic effects, the γ-receptor is postulated for sympathetic cholinergic vasodilator action in skeletal muscle

It has been widely assumed, but never proved, that the hormones secreted from the adrenal medulla markedly contribute to the vasomotor control of the blood vessels. Using a method that allowed a direct comparison of the individual actions of the hormone and vasomotor fibers on peripheral resistance in representative vascular beds, it has been found that the blood vessels are completely dominated by their vasoconstrictor fibers The contribution by the adrenal secretion is negligible, except in skeletal muscle, in which small physiological concentrations of epinephrine cause almost maximal vasodilation On the basis of adrenal catecholamine secretion following maximal excitation of the vasomotor center, it is estimated that the maximal secretion is 2 to 3 μg./kg /min Such concentrations of epinephrine and norepinephrine (probably 10 times the normal concentration), when injected intravenously, dilate the vessels of the skeletal muscles and probably those of the liver, but have insig-

nificant vasoconstrictor effects in other regions when compared with the constrictor fibers

The importance of the adrenal catecholamines for the motor control of the blood vessels is next to insignificant as compared with direct vasomotor innervation However, a moderate secretion of epinephrine induces extensive metabolic effects in liver cells (which lack direct sympathetic innervation) In skeletal muscle, physiological amounts of epinephrine produce significant β-receptor dilator response, localized chiefly to the precapillary vessels This reduction in the ratio of precapillary to postcapillary resistance, leads to filtration of fluid from the intravascular to the extravascular space Norepinephrine secretion from the adrenal gland has relatively little influence on the blood vessels, either directly or through metabolic dilation (indirectly)

RENIN-ANGIOTENSIN SYSTEM

Angiotensin is the most potent vasoconstrictor of biogenic origin known Its production following renal renin release is described elsewhere (sect 7, chap 6) The significance of the renin-angiotensin system in normal circulatory control is uncertain, and it would be premature to attribute to it a specific role It may be that angiotensin is of greater importance in the regulation of aldosterone secretion or intrarenal vascular control than as a determinant of peripheral vascular tone

VASOPRESSIN SYSTEM

Normally, plasma concentrations of the hypophysial antidiuretic hormone are too low to produce significant vasoconstrictor effects Thus, its physiological role in cardiovascular control is related primarily to long-term regulation brought about by its influence on water reabsorption in the renal tubules However, vasopressor concentrations can be reached under anesthesia and nonpressor concentrations potentiate pressor responses to catecholamines Thus, vasopressin may play a vasoconstrictor role in emergency situations such as hemorrhage and acute surgical procedures which are known to induce its release

THE ROLE OF VASCULAR TONE IN VASCULAR FUNCTIONS

Vascular tone designates the state of contractile tension in vessel walls *Basal tone* is defined as the degree of contractile tension remaining in vessels after complete elimination of known extrinsic excitatory influences Rhythmic contractions of the small precapillary vessels has been termed *vasomotion* the frequency, duration, and amplitude of these phasic constrictions in numerous arterioles in a vascular bed determine flow resistance in part and their variations represent alterations in tone The underlying smooth muscle contractile activity may be established by different mechanisms Myogenic automaticity and propagation, which may be modulated by vasomotor nerves and vasoactive agents, are characteristic of some vessels Other vessels develop tone from asynchronous spike discharges in smooth muscle cells which are dependent on impulse activity in vasomotor nerves Denervation removes much of such tone Nontetanic-graded contractures of smooth muscle may be another mechanism Possibly, these various types of contractile activity could operate in the same vessel, depending on changes in the local ionic and metabolic environment

The consecutive segments of a given vascular bed have different degrees of basal tone in precapillary resistance vessels tone is high, while it is less in capacitance vessels Flow and pressure in capillaries, which are noncontractile, is governed by contractile activity of the pre- and postcapillary vessels Thus, the number of capillaries carrying blood (i e , the active capillary surface area) is determined by the frequently rhythmic caliber variations of the precapillary sphincters It appears that capillary permeability remains a constant for a given tissue during physiological adjustments of vascular tone Capillary exchange variations primarily depend upon changes in functional capillary surface area and the ratio of precapillary and postcapillary resistance

The degree of basal tone varies in different tissues it is high in resistance vessels of the brain, myocardium, skeletal muscle, and splanchnic organs, but is practically absent in the arteriovenous anastomoses of the skin

Vascular Functions. Variations in vascular tone are determinants of vascular resistance, capillary exchange, arteriovenous shunting, vascular capacitance, and vascular compliance Luminal diameter, in accordance with the Poiseuille relation (p 3-23) is a dominant factor determining vascular resistance to flow and, thereby, tissue blood supply for a given arterial

pressure The ratio of precapillary to postcapillary resistance is a major factor influencing capillary hydrostatic pressure and thus fluid filtration and absorption in accordance with the Starling mechanism (sect 4, chap 12) The adjustment of fluid volume distributions between the intravascular and extravascular compartments depends upon net filtration or absorption which are markedly influenced by the resistance ratio of precapillary and postcapillary vessels Further, the capillary surface area available for exchange, capillary flow volume and capillary flow velocity are dependent on the resistance of precapillary sphincters Shunting of blood around capillary exchange networks is important, for example, in the skin circulation Here arteriovenous anastomoses can dilate, accommodating large volume flow for thermoregulatory purposes, without disturbing capillary flow Increased tone in capacitance vessels, venules and veins, permits mobilization of blood and decreases compliance to prevent blood pooling, factors which promote venous return to the heart Finally, alterations in arterial wall compliance influence pulsatile pressure and volume transmission through arterial circuits

VASCULAR CIRCULATORY DIMENSIONS AND VESSEL TONE

As shown in Figure 3 142, the parallel coupled vascular circuits to individual organs differ markedly in the ratio of resting to maximal flow rates Three relations are distinguishable

1) Vascular beds in which dimensions are sufficient so that with maximal dilatation blood flow and capillary surface area will meet maximal metabolic needs, e g , myocardium and central nervous system

2) Vascular beds too small to provide for maximal metabolic activity, e g , skeletal muscle, which during strenuous exercise must go into "oxygen debt "

3) Vascular beds in which maximal vascular dimensions exceed maximal metabolic needs, e g , kidney, skin, glands, and adipose tissue Here regional circulation has functions in addition to supply of nutriments and removal of wastes, e g , excretory function of the kidney, thermoregulatory function of the skin, glandular synthesis of secretory materials, and mobilization of energy stores from adipose tissue

If all circulatory beds dilated maximally at once, cardiac output could not equal the total flow rate required and pressure would fall Thus, total vascular resistance must be maintained at some minimal level by appropriate reciprocal adjustments when maximal flows prevail in some organs

SUMMARY OF REMOTE AND LOCAL CONTROL SYSTEMS

Remote control systems (e g , nervous system, adrenomedullary system, etc) function to regulate arterial blood pressure and blood volume They bring about regional and general circulatory adjustments as required to subserve such needs as thermoregulation, fluid balance, exercise, defense-alarm reactions, sexual responses, diving reflexes, etc

In muscle, vasconstrictor nerve stimulation with constant arterial and venous pressures can increase the resting 4 1 pre- to postcapillary resistance ratio two or more times The resultant lowering of capillary hydrostatic pressure, causes fluid absorption into the plasma compartment Such resistance adjustments form the basis of reflex vasomotor regulation of plasma volume called forth by arterial baro- and chemoreceptors and cardiac stretch receptors Sympathetic nerve stimulation also affects *capacitance vessels*, leading to expulsion of up to 30% of regional venous blood volume Additionally, arteriolar resistance is increased by sympathetic vasoconstrictor activity Thus, the muscle vascular bed functions importantly in reflex vasomotor adjustments of blood pressure (arteriolar resistance), plasma volume (transcapillary exchange), and venous return (capacitance vessels)

Reflex activation or direct stimulation of sympathetic adrenergic vasoconstrictors to the skin produce contraction of resistance, capacitance, and arteriovenous shunt vessels A dominant effect is constriction of resistance and arteriovenous shunt vessels, thus increasing total resistance and decreasing cutaneous flow Reflexly induced changes in cutaneous flow are important in thermoregulation, as can be demonstrated by large increases or decreases in skin flow in response to heating or cooling of diencephalic heat regulation centers or of the entire body. Reflex venoconstriction is for the most part transitory in the skin, indicating little sustained change in capacitance function Thus, skin circulatory changes are more concerned with temperature regulation, rather

than with circulatory homeostasis as in the case of skeletal muscle

In the intestine, sympathetic adrenergic activation induces initial elevation of arteriolar resistance, that is later overcome by an autoregulatory escape from the vasoconstrictor influence Changes in pre- to postcapillary resistance ratio are likewise transient, so that capillary hydrostatic pressure returns to control levels The decrease in vascular capacitance produced is sustained and can mobilize substantial volumes of blood from the splanchnic region Thus, the intestinal vascular response to increased sympathetic nerve activity is characterized by moderate alterations in arteriolar resistance and sustained decreases in blood volume with little change in capillary fluid absorption In contrast to muscle, the intestine participates little in neurogenic plasma volume control It does contribute effectively to blood pressure homeostasis through reflex arteriolar resistance and capacitance alterations

The sympathetic cholinergic vasodilator system, confined to skeletal muscle in those species in which it has been demonstrated, has its predominant effect on arteriolar vessels, while precapillary sphincters and capacitance vessels have little or no cholinergic innervation Stimulation of these fibers decreases the pre- to postcapillary resistance ratio, thus increasing net transcapillary fluid filtration The sympathetic cholinergic dilator system does not appear to be activated by homeostatic chemo- or baroreceptor reflexes Stimulation of the hypothalamic vasodilator area produces the behavioral pattern of alerting, flight, or attack in conscious animals This suggests a role for these cholinergic vasodilators in alerting and fright-flight reactions

The *parasympathetic cholinergic vasodilator fibers* function in vasodilatation associated with specific organ activity such as salivary secretion and penile erection *Dorsal root vasodilator fibers* are not activated reflexly from the central nervous system and appear to be involved only in local vascular effects of the "axon reflex" type In general, reflex vasodilatation such as that induced by baroreceptor stimulation, results from diminished sympathetic vasoconstrictor discharge rather than parasympathetic cholinergic vasodilator effects

The adrenomedullary hormones produce α-adrenergic vasoconstrictor action similar to that of sympathetic vasoconstrictor fibers The latter dominate during reflex pressor responses The β-adrenergic vasodilator effect is of unique significance since these receptors are not innervated For example, skeletal muscle β-receptor dilator response to epinephrine is significant and cannot be elicited by nerve action The metabolic actions of epinephrine, already mentioned, are also of importance in influencing local factors which govern blood flow

Local mechanisms of vascular control effect so-called autoregulation and have as their chief targets precapillary resistance vessels and sphincters Changes in the ionic and chemical metabolic milieu and vascular smooth muscle response to stretch cause vasomotor reactions appropriate to tissue metabolic needs which can oppose or override the remote effects This permits a degree of local autonomy important as a regional safeguard against the more generalized responses to homeostatic circulatory reflexes and adrenomedullary hormones Thus "metabolic vasodilators" protect against excessive vasoconstriction and myogenic control of precapillary resistance vessels and sphincters protects against capillary hydrostatic overload

The vasomotor system of nerves discussed above is the main adjustor of the peripheral circulation, whereby the blood pumped by the heart is distributed throughout the body according to the needs of each individual tissue and of the organism as a whole This vital role is accomplished by constant, moment-to-moment adjustments of the resistance in the local vascular beds, and it is mediated and regulated by the central nervous system through various reflex mechanisms (Fig 3 143)

Our knowledge and understanding of these reflex mechanisms is far from complete, the function of some of them is fairly well established, the existence of others is known without any certainty regarding their function, and it is safe to assume that there exist still others, totally unknown to us at the present time Inevitably, in the ensuing parts, the lengthiest considerations will be given to the most completely studied mechanisms, but it should be clear that this is not necessarily proportional to their respective physiological importance

Vasomotor reflexes can be elicited by the stimulation of practically any afferent nerve—somatic or visceral

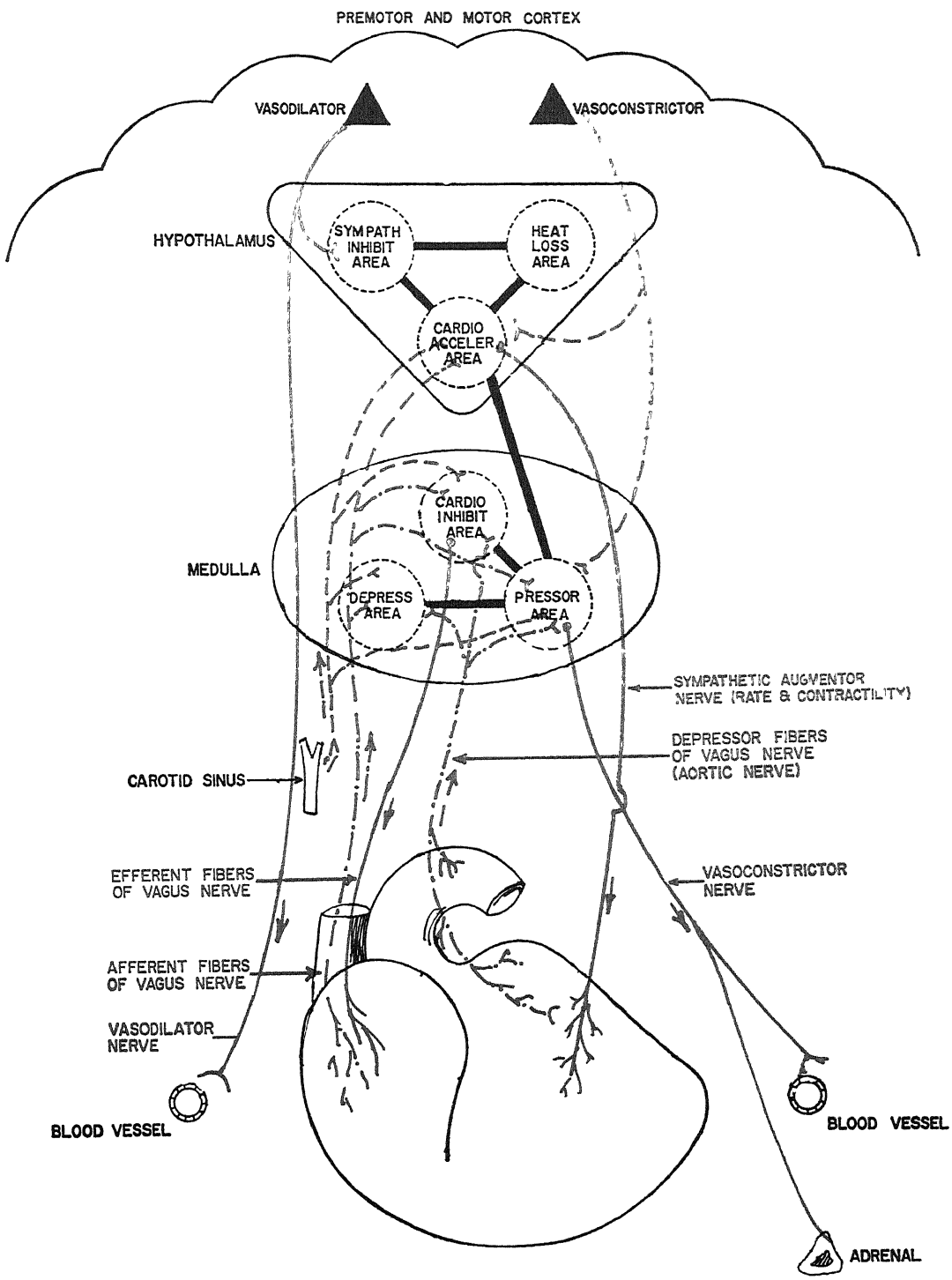


Fig 3 143 Diagrammatic representation of cardiovascular reflex mechanisms

Reflex Control Mechanisms

VASCULAR REFLEXES RESULTING FROM THE STIMULATION OF SOMATIC NERVES

Historically, these are among the earliest vascular reflex mechanisms known, although their importance in overall circulatory regulation is relatively minor

Stimulation of the central end of a nerve, such as the sciatic, the median or a sensory cranial nerve, may result in either a rise or a fall in the arterial blood pressure according to the strength and type of the stimulus employed The components of the reflex arc upon which the responses depend are *1*) afferent fibers in the peripheral nerve, *2*) the vasomotor centers, and *3*) the efferent vascular nerves, i e , the vasoconstrictors or vasodilators To elicit a pressor reflex, a stimulus much stronger than that necessary to provoke a depressor response must, as a rule, be applied, that is, one which would elicit pain in a conscious animal In the elicitation of either reflex, the magnitude of the response is apparently dependent upon the number of afferent fibers involved For example, stimulation of various nerves of the brachial or lumbar plexus causes practically equivalent depressions or elevations in the blood pressure when the number of afferent fibers in the respective nerves is taken into account

It must be re-emphasized that not all the pressor or depressor impulses ascend to the medulla, but that reflex arcs exist which have their centers within the cord, both vasoconstrictor and vasodilator reflexes can be elicited in an animal whose cord has been divided in the lower cervical region a short time previously

From the results of experiments upon animals, it is to be expected that in the human subject a painful stimulus applied to a somatic nerve will be followed by a pressor response The excitation of psychic centers and, also, the liberation of epinephrine are additional factors which play an important part in the pressor response resulting from painful stimuli On the other hand, there are numerous afferent fibers in the hollow viscera which respond to stretch and result in a decreased blood pressure Stimulation of the mesentery, peritoneum, and abdominal viscera or of certain regions such as the anus, vagina, and spermatic cord may be followed by a fall in blood pressure Finally, there may be obvious reflex effects which may or may not be associated with blood pressure changes For example, in normal persons with bladder distention, there may be facial flushing, but with evidence of widespread venous constriction Following inflation of balloons placed in the esophagus, rectum, or in an ileostomy, regional venoconstriction also can be demonstrated

ARTERIAL STRETCH RECEPTORS

Experiments have shown that the fundamental regulation of systemic blood pressure is reflexly controlled by the action of the arterial pressure itself on stretch sensitive receptors (also termed *baroreceptors* or *pressoreceptors*) located in the vascular walls, especially of the aortic arch and carotid sinus areas

Aortic or Cardiac Depressor Nerve The branch of the vagus which is known as the *aortic* or *cardiac depressor* nerve is purely afferent and depressor in function, when it is sectioned, and its central (cerebral) end stimulated, a pronounced fall in pressure occurs (Fig 3 144), excitation of the cardiac end, on the other hand, causes no effect Two factors are involved in the depressor response following stimulation of the fibers

Slowing of the Heart Rate The efferent fibers of the vagus of the same and of the opposite side constitute the efferent limb of the reflex arc through which this response is chiefly brought about, for its full elicitation at least one vagus must, therefore, remain intact

Vasodilation The vasomotor pathways constitute the efferent limb of this reflex Vasoconstriction is inhibited and vasodilation results

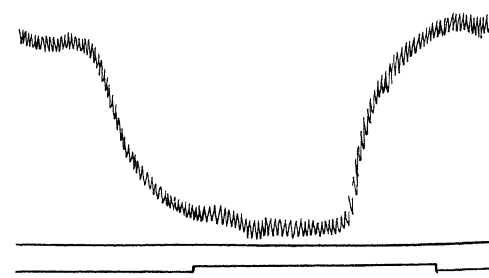


Fig 3 144 Fall in arterial blood pressure resulting from stimulation of the central end of the cardiac depressor (aortic) nerve The drum was stopped in the middle of the curve and the excitation maintained for 17 min The line of zero pressure should be 30 mm lower than here shown (From Bayliss 1893)

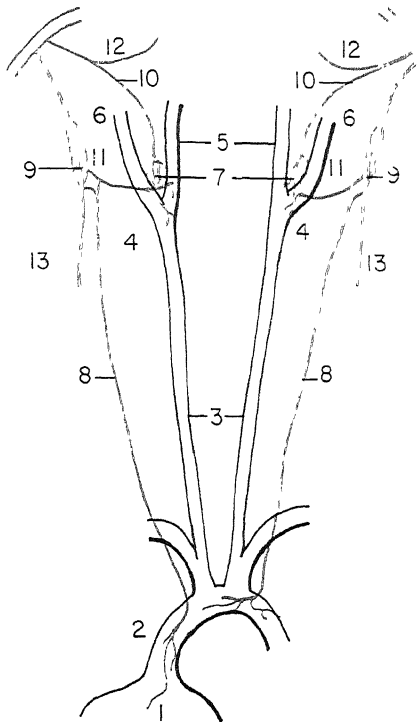


Fig 3 145 Innervation of the carotid sinus and arch of aorta *1* heart *2* arch of aorta *3* common carotid *4* carotid sinus *5* external carotid *6* internal carotid *7* carotid bodies *8* cardiac depressor nerve *9* ganglion of vagus *10* sinus nerve branch of the glossopharyngeal nerve *11* nerve branch connecting the carotid sinus with the vagus ganglion *12* glossopharyngeal nerve and *13* vagus nerve (After Heymans et al 1933)

The reflex effect upon the vessels cannot, therefore, be elicited after section of the spinal cord in the lower cervical region

The receptors of the reflex (the terminals of the aortic nerve) are situated in the aortic arch and upper part of the thoracic aorta, in the ventricles and possibly also, according to Daly and Verney (1926), in the coronary and pulmonary vessels (see Fig 3 145)

The fall in pressure is due mainly to dilation of the splanchnic vessels The dilation is not, however, confined to these vessels but includes those of the skin and muscles Cardiac slowing plays a role in the production of the fall in blood pressure, although almost as great an effect can be obtained after both vagi have been cut The depressor reflex can be elicited by mechanical or electrical stimulation of the aortic wall itself, wherein the special proprioceptors are located, stretching is an especially effective type of stimulus It was shown that action currents ascend the nerve synchronously with the heart beats (Fig 3 146) The normal stimulus is evidently the pulsatile expansion of the aortic wall, a rise in general blood pressure increasing the intensity of the stimulus a fall in pressure causing the reverse effect The primary reflex response of the pulmonary vascular bed to carotid sinus and aortic arch pressoreceptor stimulation is vasodilatation, a response which may be masked by bronchial artery dilatation and redistribution of blood from bronchial to pulmonary circuits (Daly and Hebb, 1966)

Carotid Sinus Mechanism. The carotid sinus is the term applied to the slight enlargement of the common carotid artery where it bifurcates into the internal and external carotids (Fig 3 145) The carotid sinus was shown by Hering in 1923 to play an important role in the regulation of the cardiac rate and arterial blood pressure Compression of the carotid at its bifurcation (so as to raise the pressure within the sinus) causes a marked slowing of the heart rate, vasodilation (see p 3-179), and

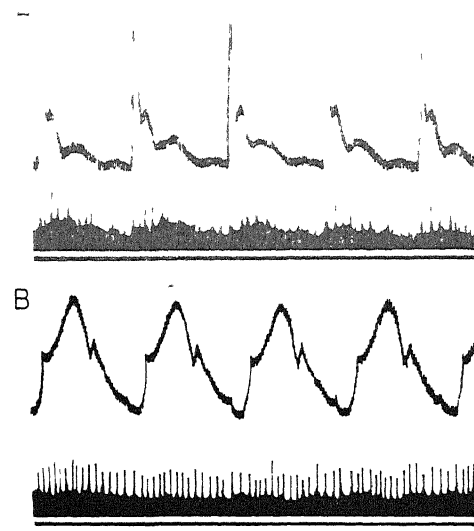


Fig 3 146 The upper curve in each record represents the arterial blood pressure registered by a membrane manometer the lower curve of each record shows the electrical discharge from a single fiber of the carotid sinus nerve of the rabbit In the upper record (*A*) the mean arterial pressure was 55 mm Hg even at this low level a discharge of 4 impulses accompanied each ventricular systole In the lower record (*B*) the mean arterial pressure was 135 mm Hg in this instance there was a more rapid and more continuous discharge from the end-organ (Bronk and Stella 1932)

a fall in blood pressure, these effects result even though mechanical stimulation of the vagus is carefully avoided Electrical stimulation of the sinus wall produces similar effects Pressure upon the common carotid some distance below the sinus (so as to reduce the blood pressure within the sinus itself) causes cardiac acceleration, vasoconstriction, and a rise in arterial pressure together with, as shown by Heymans, the liberation of epinephrine The carotid sinus therefore constitutes a mechanism whereby both pressor and depressor effects are mediated The effects are brought about through the following neural mechanism

The Sinus Reflex Arc. The afferent fibers of the reflex arc are contained in the *sinus nerve*, a branch of the glossopharyngeal This delicate filament descends between the internal and external carotids to the sinus where its fibers terminate in sensory organs (proprioceptors) situated between the connective tissue fibers in the adventitia of the sinus wall Centrally the fibers of the sinus nerve make connections with the cardioinhibitory and vasomotor centers The efferent limb of the cardiac part of the reflex is the vagus The efferent limbs of the vasodilator and vasoconstrictor reflexes are apparently sympathetic fibers, for these reflexes are abolished by complete removal of the sympathetic chains

Study of the action currents passing along the sinus nerve shows that impulses stimulated by the arterial pressure pulse are discharged throughout the cardiac cycle, their frequency increasing during systole and decreasing during diastole A rise in general blood pressure increases the rate of impulse discharge as well as the number of sense organs excited The latter show slow adaptation so that though the stimulus (distension of the arterial wall) persists the impulse discharge shows little reduction in frequency, and when the pressure is very high, they extend with little reduction in rate throughout diastole At low pressures, impulses are discharged only during systole

The sinus reflexes have been studied exhaustively by Heymans and his associates (1958) They carried out cross-circulation experiments which speak conclusively for the physiological importance of these reflexes in cardiovascular regulation (see Fig 3 147) The sinus of one dog (B) was isolated from the general circulation and perfused with the blood of another animal (A) in the manner shown in the figure

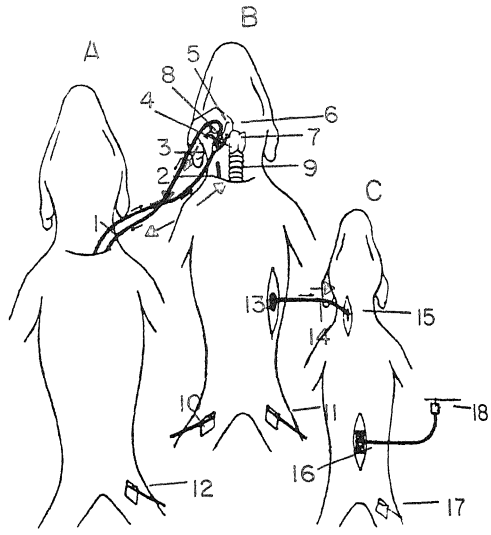


Fig 3 147 Scheme of perfusion of the isolated carotid sinus of dog *B* by dog *A* and an anastomosis between the suprarenal vein of *B* and the jugular vein of dog *C* *1* left carotid artery of dog *A* *2* right carotid artery of *B* anastomosed with carotid of *A* *3* left external jugular vein of *A* *4* isolated right carotid sinus of *B* *5* lingual artery of *B* anastomosed with jugular vein of dog *A*, *6* nerve supply to carotid sinus of *B* The blood from dog *A* flows through the carotid sinus of dog *B* and back to *A* via the lingual artery of *B* and the external jugular of *A* *7* internal carotid *8* facial and maxillary arteries *9* common carotid *10* *11* *12* and *17* femoral arteries to manometers *13* adrenal gland *14* and *15* suprarenal-jugular anastomosis *16* decapsulated spleen in plethysmograph *18* piston recorder for plethysmograph (After Heymans et al., 1933)

The nerve supply to the sinus was left intact When the arterial pressure of dog A was raised, that of dog B, recorded in the femoral artery, fell Conversely, a reduction in blood pressure of dog A caused a rise in the blood pressure of dog B In the latter instance, epinephrine liberation also occurred which was a contributory factor in the blood pressure elevation as indicated by its pressor effect upon the circulation of dog C, connected to the suprarenal vein of B These effects could not be obtained after denervation of the sinus

The sinus and aortic nerves, or so-called "buffer" nerves, constitute a mechanism of the utmost importance in controlling the arterial blood pressure and in maintaining the circulation to the brain The rise in diastolic pressure and the increase in heart rate which occur when the body changes from the recumbent to the sitting position, or from the sitting to the standing position are apparently brought about

through these nerves, they therefore play an essential part in compensating for the effect of gravity upon the circulation

An underfilled state of the vessels, as may result from hemorrhage or shock, or any other condition which tends to cause a fall in blood pressure, will call these mechanisms into play A generalized vasoconstriction results to adjust the vascular capacity to the reduced blood volume and thus maintain the blood pressure Excessive elevation of the blood pressure, on the other hand, is countered by a depressor reflex (see diagram Fig 3 143) The great importance of these reflex mechanisms in hemorrhage is shown by the fact that in an animal in which all four buffer nerves have been sectioned the rapid loss of only about one-tenth of the blood volume proves fatal, whereas usually a reduction in blood volume of from 35 to 45% is required to cause death Mayerson (1942) found that tilting anesthetized dogs from the horizontal to the upright position caused a sharp drop in blood pressure, followed within 10 sec by a compensatory rise After section of both sets of buffer nerves the compensatory rise did not, as a rule, occur

Moderate arterial hypertension can be produced in dogs by denervation of the sinoaortic baroreceptors (Heymans and Neil, 1958) It is characterized by increased lability of arterial pressure with marked hypertensive response to various external stimuli and, in one series (Ferrario et al, 1969), a paradoxical increase in blood pressure during sleep

A rise of pressure within the carotid sinus produces the behavioral signs of sleep (Koch, 1932), can diminish or abolish drug induced convulsions, and induce cortical electroencephalographic synchronization The physiological significance of this effect in the regulation of sleep-wake patterns is somewhat uncertain Studies in cats suggest that synchronous inflow from baroreceptors is in part responsible for synchronous sleep (Baust and Heinemann, 1967)

Mechanism of Action. The mechanism of action of the blood pressure on the baroreceptors is complex and only partially understood They are essentially stretch receptors and if deformation of the baroreceptive arterial walls from a raised arterial pressure is prevented, the baroreceptors no longer respond to the pressure change Reflex blood pressure depression from carotid sinus stimulation is greater with pulsatile than with steady pressure stimulation At moderate sinusoidal stimulating pressures within the carotid sinus the blood pressure depression increases with both frequency and amplitude of the stimulating pressure This frequency effect disappears at higher intrasinal pressures and if stimulation pressure exceeds a range of 160 to 180 mm Hg (in dogs) the blood pressure reaction is inverted (Stegemann and Tibes, 1969)

Local application of drugs, such as epinephrine or norepinephrine, increases the smooth muscle tension in the arterial wall of the isolated carotid sinus preparation which, in turn, is associated with a large increase in impulse traffic along the baroreceptor fibers throughout the cardiac cycle and a resetting of the baroreceptive mechanisms at a higher level, with resultant blood pressure adjustment Sodium nitrite, which relaxes smooth muscle, reduces the intramural tension in these sensitive regions decreases the impulse traffic, and resets the blood pressure regulation at a lower level

In experimental and clinical hypertension the pressure sensitivity of the carotid sinus receptor response is reduced, thus baroreceptor response can be "reset" in its range of function to maintain blood pressure at a higher level When hypertension is relieved the range of function can be "reset" downward (McCubbin et al, 1968) These observations indicate that under normal conditions, the response of the baroreceptor area and the level of blood pressure maintained depend not only on the level of arterial pressure, but, possibly, even more so on the resistance of the baroreceptive arterial walls to deformation by the intravascular pressure

Other Arterial Baroreceptors. Possible additional arterial baroreceptor types have been studied in the common carotid artery, the thoracic, and mesenteric arteries Experiments in which a sudden rise in intracranial pressure causes systemic hypertension, and sudden reduction in intracranial pressure leads to systemic hypotension, suggest, but do not prove, that pressure sensitive receptors are present in the cranial cavity. However, the systemic hypertensive effect of increased intracranial pressure (*Cushing effect*) is largely, or entirely, secondary to asphyxia of medullary centers (see pp 3-208 and 3-211) Cushing (1902) first demonstrated in animals that an acute rise in intracranial pressure induces an increase in systemic arterial pressure which just exceeds

that within the cranial cavity Impulse output in the mesenteric baroreceptors exists throughout the cardiac cycle, and, if pressure is changed in the mesenteric artery or the artery is clamped, vasomotor reflexes are induced These reflexes do affect the systemic blood pressure even when the sinoaortic nerves are cut It is a segmental, reflex adaptation of the circulation and vasomotor state related to pressure changes in a given area It apparently exists in the cat but not in the dog Reflex responses can also be obtained in the innervated perfused hind leg of the open-chest dog, when the descending aorta between the left subclavian artery and the diaphragm is isolated and perfused at varying pulsatile pressures, or when epinephrine is added

The importance of these afferent mechanisms in control of the circulation is not known

BALANCED AND RECIPROCAL VASCULAR REACTIONS

In the intact animal the height of the blood pressure at any moment, insofar as the nervous control of the peripheral vessels is concerned, is apparently the algebraic sum of the effects of afferent impulses impinging upon the vasomotor centers Under ordinary circumstances, impulses arising from the carotid sinus and aortic arch play the most prominent roles, but impulses from skin, muscles, and viscera and from higher nervous centers also exert an important influence That pronounced effects upon peripheral vessels can be produced by radiation of impulses from higher centers is evidenced by such phenomena as blushing, pallor, penile erection, and certain types of syncope (fainting) When a dog is excited, similar vascular changes can be observed in the mucosa of the exteriorized colon and in splenic volume Even very mild excitation of psychic centers exerts an influence upon the vascular mechanisms The psychogalvanic reflex (due to changes in the electrical resistance of the skin) has a vascular basis A reciprocal relationship also exists between splanchnic and cutaneous vascular areas on the one hand and the vessels of the muscles on the other Epinephrine, for example, causes dilation of the latter vessels accompanied by vasoconstriction in the skin and abdominal viscera Stimulation of the wall of a large vein or distension of the duodenum causes reflex constriction of the cutaneous vessels, and stimulation of the skin results in an increased amount of blood in the liver and in the renal cortex Furthermore, the muscular and cutaneous tissues may show opposite vascular reactions Thus, in the dog, cooling of the body causes constriction of the skin vessels and vasodilation in the muscles

A mechanism also exists in the experimental animal, whereby afferent impulses arising within an organ, although leading to generalized vasoconstriction, result locally in dilation and increased blood flow This is sometimes called the *Loven reflex* after the investigator who first described it in 1866

An interesting example of the complex nature of the reflex vascular adjustments has been described An electrical record obtained from a nerve twig coming from a Pacinian corpuscle in the mesentery during perfusion of the mesenteric vessels showed an increase in the frequency of the afferent impulses when the perfusion pressure was raised The impulse frequency was reduced by bleeding the animal and increased again when the blood was reintroduced into the body It is suggested that the Pacinian corpuscles, which lie in close relation to the vessels, are stimulated when the latter dilate Messages pass to the vasomotor center which then discharges vasoconstrictor impulses to the vessels of the area from which the afferent impulses arose Thus, undue distention of the vessels and pooling of blood in the splanchnic region is prevented It has also been shown that stimulation of the peripheral end of the splanchnic nerve causes a much greater rise in blood pressure than usual if the carotid sinuses have been excluded from the circulation (by clamping the carotids) This observation indicates that ordinarily the pressor effect of splanchnic stimulation is largely counteracted by a depressor reflex initiated from the sinus

EXTRARESISTANCE (CARDIAC) EFFECTS OF BARORECEPTOR REFLEXES

Most important changes during reflex, systemic hypertension of sinoaortic origin are *venoconstriction* (effects on the capacity vessels) and direct sympathetic stimulation of the heart Occlusion of the common carotid arteries causes systemic hypertension with resultant decrease in the plethysmographic volume of spleen, limb, kidney, and other parts of the circulation If the systemic hypertension is ascribed to widespread arteriolar vasocon-

striction, cardiac output would be expected to fall or, at least, not to increase However, total pulmonary, muscle and kidney blood flow have been shown to increase during carotid occlusion There is always some increase in peripheral resistance but it is of less quantitative significance, since the cardiac output increases Experiments, utilizing an extracorporeal circulation with exclusion of the heart and lungs to allow continuous measurement of alterations of intravascular blood volume and of venous return, have also demonstrated venoconstriction (decreased vascular volume) and an increased venous return to the heart when hypotension was induced in the isolated canine carotid sinus The opposite effects occurred with carotid sinus hypertension Venoconstriction has also been shown in the forelimb vascular bed of the dog during carotid baroreceptor stimulation Some studies have failed to confirm these convincing data

Reflex sympathetic and parasympathetic effects can significantly alter the atrial and ventricular contractility The circulatory response to carotid sinus hypotension can be, dominantly, an increased cardiac action and, secondarily, an increase in peripheral resistance

In unanesthetized dogs electrical stimulation of the carotid sinus nerves produced vagal cardiac slowing, arterial hypotension, and decreased mesenteric, renal, and hind limb vascular resistance (Vatner et al , 1970) The most profound decrease in resistance and increase in flow occurred in the hind limb The increase in hind limb flow took place also in exercising animals when the muscular bed was already dilated on a metabolic basis Observations in man are instructive (Fig 3 148) Pressures within the carotid sinus regions were directly measured with indwelling catheters in conscious, supine man while the internal pressure was lowered by central, bilateral, carotid artery compression As the carotid sinus pressure dropped from a mean of 100 to 40 mm Hg, the arterial blood pressure and heart rate increased only mildly There is no evidence that the pressor response is secondary to reduction in caliber of the limb and skeletal muscle vessels The resistance to flow which they offer is actually slightly less than in the control state as the result of a mild increase in systemic blood pressure and in blood flow through the limbs Simultaneous measurement of cardiac output

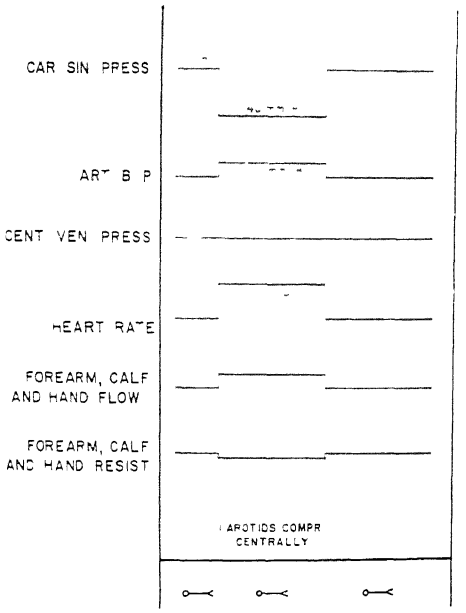


Fig 3 148 Bilateral carotid artery compression in supine man producing carotid sinus hypotension reflexly causes a rise in systemic pressure a tachycardia and an increase in peripheral flow with concomitant decrease in peripheral resistance

would have been helpful here, since the increase in arterial blood pressure could arise from the increased cardiac output or from increased vessel resistance in another area

CARDIOVASCULAR REFLEXES OF CHEMORECEPTOR ORIGIN

The carotid and aortic bodies adjacent to the carotid sinus and root of the aorta contain epitheloid cells and nerve endings sensitive to the chemical composition of the arterial blood The fibers of the aortic body run in the vagus, while the fibers of the carotid body are branches of the glossopharyngeal nerves The aortic body is connected with a branch of a fine artery arising from the aorta beyond its arch The carotid body (glomus caroticum) is a small structure situated upon a branch of the occipital artery, or upon a small vessel arising directly from the external carotid just above the bifurcation of the common carotid It is composed of rounded clumps of polyhedral cells, and possesses a rich network of capillaries of sinusoidal character Both the carotid and aortic chemoreceptors sample the blood for its Po_2, Pco_2, pH, and possibly for other qualities Blood flow through the carotid body appears to be by far the

largest for any body tissue Calculations based on direct flow measurements indicate a flow of 2000 ml /100 g /min (left ventricle = 80 to 100 ml /100 g /min) while the oxygen usage of 9 ml /100 g /min , compares with that of the left ventricle This flow can be cut to one-third by stimulation of the cervical sympathetic nerves Reflex vasomotor effects evoked by chemoreceptor stimulation have been extensively studied In general, the chemoreceptors are stimulated by anoxia, hypercapnia, and acidosis The mechanism by which changes in blood chemical composition effect a response in these two structures is not clearly understood nor has the chemoreceptor intermediate that stimulates the nerve endings been identified, but it has been shown that electroneurographic evidence of chemoreceptor activity exists at CO_2 tensions above 30 mm Hg and at O_2 tensions of approximately 90 mm Hg or less With increasing Pco_2 and decreasing Po_2, discharge of the chemoreceptors is diffuse, but the discharge with increasing CO_2 tension is rarely as great as that resulting from anoxia Acute hypoxia (inhalation of low oxygen mixtures or of nitrogen) causes systemic hypertension and vasoconstriction (decreased blood flow) in the limbs and intestine, responses are abolished if the chemoreceptor areas are blocked It is unlikely that chemoreflexes cause systemic hypertension only by inducing an increase in arteriolar resistance, but there are no studies of the reflex effects of such stimulation on cardiac output or on the capacitance vessels It is certain, however, that they contribute to the maintenance of the circulation following depletion of the blood volume The latter can be shown in an animal in which, following massive hemorrhage, the blood pressure decreased to 70 mm Hg There was intense firing of the chemoreceptors which could be stopped by raising the oxygen tension while resection of the sinus nerves caused a further fall of blood pressure The latter can only be due to withdrawal of chemoreceptor reflex vasoconstriction, since the baroreceptors are not operative at this blood pressure level Presumably the capacitance vessels (veins) are involved here as well as the resistance vessels

VASCULAR EFFECTS OF NORMAL VARIATIONS IN BLOOD OXYGEN, CARBON DIOXIDE AND pH

Cardiac response to chemoreceptor stimulation is complex In experimental preparations arranged so as to prevent secondary influence of respiratory stimulation on cardiac activity, stimulation of the carotid bodies produces reflex bradycardia while stimulation of the aortic bodies produces reflex tachycardia Where respiratory effects are allowed, cardiac response to carotid body stimulation is related to the intensity of the ventilatory augmentation With mild ventilatory stimulation the heart rate is usually decreased, with marked increase in ventilation, cardiac acceleration occurs (Berne and Levy, 1967)

The primary effect of chemoreceptor stimulation on the pulmonary circulation is vasoconstriction, which may be masked, however, by the vasomotor changes in the bronchial (systemic) circulation (Daly and Hebb, 1966)

It is generally stated that both carbon dioxide and oxygen have dual effects Although CO_2 may have a minor reflex effect through the aortic and carotid body chemoreceptors, it may also have a large direct central effect on the medullary centers which leads to peripheral vasoconstriction with rising blood CO_2 concentrations, while with decreasing blood CO_2, the center will ultimately cease giving out its normal tonic influence It also may have a large direct peripheral vessel effect, an increasing blood CO_2 causing peripheral dilation and a decreasing blood CO_2 leading to vasoconstriction Similarly, the central effects of oxygen lack are supposedly largely mediated through the sinus and aortic bodies, a decrease in oxygen leading to reflex peripheral vasoconstriction, while in the periphery, local hypoxia has the effect of vasodilation While these statements based on experimental data have been perpetuated, it is very difficult to make such a functional separation and yet retain some reasonable degree of normalcy in the experimental approach to the problem Aside from the effects mediated through the chemoreceptors, it is doubtful that the relative effects of physiological variations in blood gases and pH on the vasomotor centers and the general peripheral circulation are known However, for some individual organs or regions, one can distinguish between the local vascular effects of chemical changes in the blood and their central neural effects For example, the main body of evidence indicates that small increases in carbon dioxide have a direct vasodilating action on cerebral and skin vessels in animals and in humans

The complexity of the problem, even in indi-

vidual regions, is indicated when one compares the effect of carbon dioxide on the vasomotor state in the extremities The hind limb of an animal perfused with blood from a heart-lung preparation shows vasodilation (greater blood flow at the same perfusion pressure) when the lungs are ventilated by air containing a small amount of carbon dioxide or by air with some reduction in its oxygen content The vasodilation must result from local chemical reactions However, in human subjects with hypercapnia alone, induced from breathing a small concentration of carbon dioxide in the inspired air, the forearm's vascular resistance (mainly muscle) is greatly increased, for systemic blood pressure is increased with little or no change in local blood flow In the same subjects in whom hypoxia was produced by breathing 5 to 10% oxygen in nitrogen, forearm blood flow increases 100% in the presence of a constant systemic arterial and venous pressure so that vascular resistance in the forearm is considerably reduced Since the same results are observed in the nerve-blocked arm, the local change in vasomotor state is due to a humoral mechanism and not to a nervous mechanism When the increased elimination of CO_2 during hypoxia is prevented by adding enough CO_2 to the oxygen-poor gas mixture to maintain the alveolar P_{CO_2} constant, vascular resistance in the forearm is not significantly changed This indicates that marked hypoxia does not affect local overall vessel resistance Thus, in man, a humoral vasoconstriction occurs with hypercapnia, a humoral vasodilation occurs with hypoxia in the muscles of the forearm which is due to the hypocapnia resulting from the hypoxia rather than directly to the lack of oxygen The exact mechanism is not clear The CO_2 or altered pH could have had a local effect, but it is just as probable that a rapid change in arterial P_{CO_2} evoked biochemical changes in the blood which caused local vasoconstriction or dilation Of several possible humoral vasodilators which are known, epinephrine, which causes vasodilation in skeletal muscle, might have been released by the lowered P_{CO_2}

In the past, it has been debatable whether the circulatory effects observed with respiratory or metabolic acidosis come from the elevated P_{CO_2}, the fall in pH, or a combination of both, because there has not been any good way of separating these effects However, with the use of the organic buffer Tris (hydroxymethyl) amino methane, reasonable experimental separation has been performed with the finding that the cardiovascular effects arise from the decrease in pH and not from the associated change in blood Pa_{CO_2} (Fig 3 149) To make this study, the effects of metabolic acidosis were observed in dogs under apneic oxygenation with and without Tris injection Apneic oxygenation is a condition of ventilatory arrest during which oxygenation of the blood is maintained while acute hypercapnia develops This condition is produced by ventilating the dog with 100% oxygen in an open circuit for 1 hr During this time, oxygen replaces nitrogen throughout the body After this period of denitrogenation, apnea is induced by succinylcholine, and the lungs are left connected to a reservoir of oxygen Thus, the blood is maintained

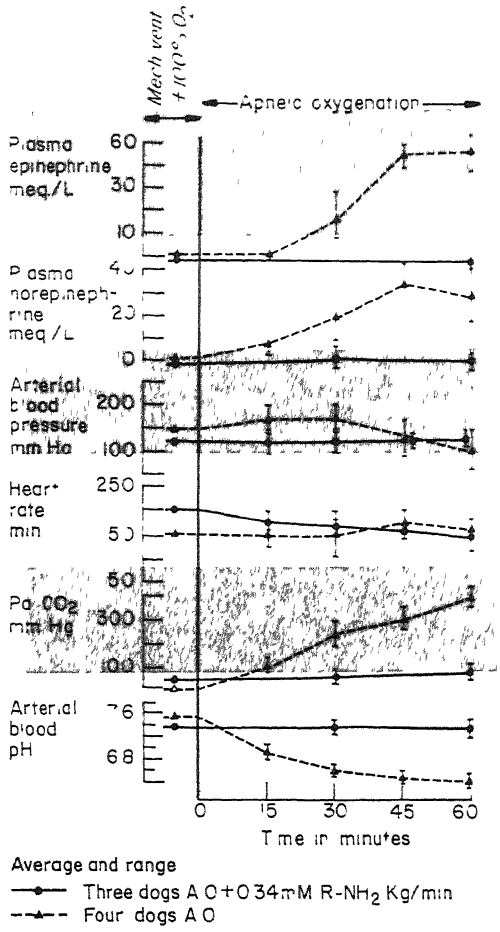


Fig 3 149 Comparative changes in pH P_{CO_2} blood pressure and catecholamine levels during apneic oxygenation with and without tris administration (Nahas et al 1959)

in full oxygen saturation, but blood CO_2 accumulates slowly Without the presence of Tris in such dogs, there is a rise in CO_2, arterial, pulmonary artery and central venous blood pressures, Pa_{CO_2}, and plasma norepinephrine and epinephrine, the blood pH drops precipitously, the heart rate mildly In similar dogs treated with intravenous drip of Tris, all these physiological variables remain close to their control values except Pa_{CO_2} which rises from 38 to 82 mm Hg Since such a P_{CO_2} increase in the intact animal does produce the above circulatory changes when it is accompanied by a parallel fall in pH, the H^+ concentration increase during hypercapnia seems to be a determinant factor in the stimulation of the cardiovascular system This type of experiment, however, still does not clearly separate the effects of CO_2 and pH, for the latter does change (7 45 to 7 31), and the vascular system may be very sensitive to it

REFLEXES FROM THE HEART AND LUNGS

From various intrathoracic vessels (including the chambers of the heart) arise other reflexes affecting vasomotor tone Although their importance in the regulation of the circulation may possibly be comparable to that of the previously mentioned reflexes, much less is known about their function, since it is much more difficult to investigate cardiopulmonary reflexes than those related to the baroreceptors in the systemic circuit

Right Heart Reflexes. Observations have shown conclusively that in the walls of the atria and great veins, there are receptors which have fairly large afferent nerve fibers and which respond to change of pressure in the atria by firing during atrial contraction (type A), or during the passive distention of atrial diastole (type B) These afferent fibers are in the vagi

In 1914, Bainbridge showed that the intravenous injection of saline or blood produced tachycardia in the anesthetized dog and that this was abolished only by bilateral vagotomy The location and nature of receptors responsible for this reflex have not been identified and fluid infusions frequently fail to elicit tachycardia It has been shown in anesthetized dogs that intravenous infusions increase the heart rate when the initial rate is below 140/min, but actually decrease the heart rate at more rapid initial rates Thus both the mechanism and physiological role of this reflex remain controversial.

In anesthetized dogs, partial occlusion of the pulmonary artery leads to an increase in pulmonary blood flow It has been shown that this reflex is dependent on an intact sympathetic innervation of the heart, the vagi and medullary centers are not involved Presumably the increase in pulmonary blood flow is due to a reflex myocardial stimulation initiated by a rise in right ventricular systolic pressure and is mediated by the cardiac sympathetic nerves

Pulmonary Vessel Reflexes Most investigators now admit that the pulmonary blood vessels have tone and, therefore, the possibility that this tone can be altered Daly has found evidence in the anesthetized dog that the pulmonary vascular resistance can be altered by extrinsic vasomotor nerves Both vasoconstrictor and vasodilator responses have been demonstrated The primary effect of chemoreceptor (carotid body) stimulation is pulmonary vasoconstriction, although the reflex vasoconstrictor effect on the bronchial circulation, which occurs simultaneously, may actually produce a decrease in pulmonary vascular resistance (Daly and Hebb, 1966)

Pulmonary Artery Reflexes Elevation of pressure in the main pulmonary artery and its bifurcation within physiological limits reflexly induces systemic hypotension and inconstant bradycardia Vagotomy abolishes the response Elevating the pressure above physiological limits (e g, 80 to 120 mm Hg) causes, paradoxically, a rise in systemic arterial pressure

Reflexes from the Smaller Pulmonary Vessels Despite a tremendous amount of investigation, it is problematic whether reflexes are initiated from the smaller vessels of the pulmonary circuit—arterioles, capillaries, and venules Embolization by clots, starch grain, glass, and plastic beads had not determined conclusively whether the results have a mechanical or a reflex explanation From the work of Daly (see Daly and Hebb, 1966), it is known that the sympathetic innervation of the vessels is largely vasoconstrictor and that stimulation of the stellates may, at times, reduce lung blood flow by as much as 30% A variety of experiments have demonstrated pulmonary vasoconstriction resulting from such stimuli as pulmonary embolization and venous congestion There are indications that some of these responses may be mediated by local pulmonary axon reflexes

An "alveolar-vascular reflex of the lung" has been demonstrated If water is introduced into the airways of sheep, a reflex response occurs

via the parasympathetic nervous system (blocked by atropine) producing pulmonary vasoconstriction The vagus is not involved for vagotomy does not abolish the response, the reflex is therefore considered to be local in nature This reflex would lead to a homeostatic redistribution of blood flow away from nonventilated lung areas (see also p 3-210)

Left Heart Reflexes. *Left Ventricle* Daly and Verney (1926) were first to show that pressor receptors existed in the left heart In an innervated heart-lung preparation in which the aortic pressure was kept constant, an increased pressure in the left side of the heart caused cardiac slowing More recently, it has been shown that when the pressure is elevated in a vascularly isolated left heart, reflex bradycardia, systemic hypotension, and vasodilation of the leg vessels result Since left atrial pressure elevation alone was ineffective, it is concluded the receptors were in the left ventricle Upon vagal blockage the effect disappeared These effects may also arise from pressoreceptors of the aortic arch which are known to extend into the left side of the heart

Ross and co-workers (1961a, b) studied this reflex by perfusing the systemic vascular bed, using an extracorporeal circulation, pressures could be varied independently in the innervated but isolated heart Significant decreases in systemic vascular resistance occurred when left ventricular systolic pressure was elevated, or when left ventricular diastolic and mean left atrial pressure were *also* elevated Only minimal alterations in vascular resistance were observed to accompany right ventricular pressure elevations alone Most important, the elevations in left intracardiac pressure produced a large increase in systemic vascular volume and a diminished venous return This suggests that left ventricular baroreceptor stimulation reflexly inhibits venous tone Informative experiments are those of Doutheil and Kramer (1959), who varied the left ventricular stretch and pressure by an adjustable clamp on the aorta, while the systemic flow was metered, and its perfusion pressure maintained constant during aortic constriction by a very large pressure buffering chamber attached to the aorta It was observed (Fig 3 150) that within 4 sec after elevation of left ventricular pressure by mechanical constriction, the aortic flow starts to rise and is almost doubled, indicating massive peripheral dilation arising reflexly from the beginning of the aorta or in the left heart The

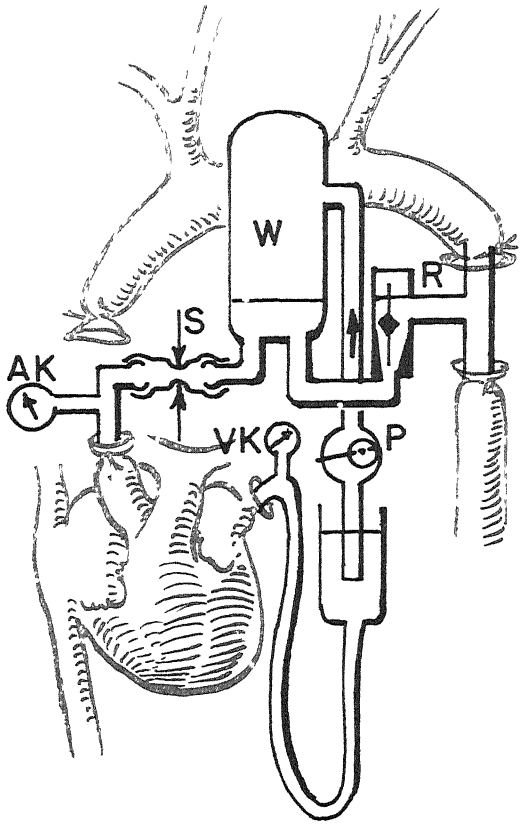


Fig 3 150 Experimental device to show presence of left ventricular pressure (stretch) receptors by aortic compression *S* clamp *AK* and *VK* aortic and left atrial pressure manometers respectively *W* constant pressure buffering chamber *R* systemic flow meter *P* pump (Doutheil and Kramer 1959)

afferent pathway for this reflex was demonstrated to be the vagus nerve, for following its cervical section, the effect disappeared To what extent this reflex is invoked under normal conditions is problematic, since a rather large increase in ventricular pressure (and presumably, therefore, stretch) is necessary to elicit a systemic flow response

Coronary Chemoreflex (Bezold-Jarisch Reflex) Intravenous injection of veratridine and certain other agents (e g , nicotine, some antihistaminics, etc) produces reflex cardiac slowing, hypotension, and apnea (the *Bezold-Jarisch* reflex) The reflex apnea is caused by excitation of the receptors for the Hering-Breuer (lung inflation) reflex The receptors responsible for the bradycardia and hypotension can be reached through injections into the coronary arteries or the pericardial sac In the cat, injections into the right or left coronary

arteries elicit the response while in the dog only injections into the left coronary arteries or pericardial sac are effective (Comroe, 1965, Aviado, 1965) The hypotension is independent of the bradycardia, being caused by reflex vasodilation

Left Atrial Receptors Receptors in the left atrium (and probably pulmonary veins) which respond to stretch have been implicated in the regulation of blood volume (Herd, 1970, Gauer et al , 1970) Acute distention of the left atrium, as with a balloon, increases heart rate, in general decreases arterial pressure (although the reverse may occur), and causes diuresis Procedures which produce a shift of blood into the intrathoracic circulation, thus elevating central blood volume and atrial pressure, produce diuresis and lead to reduction of plasma volume This redistribution of blood volume can be produced, for example, by continuous *negative pressure breathing*, head-out *body immersion* in a thermoindifferent bath, and human centrifuge simulated *gravitational effects* (so-called transverse G) The reduction in plasma volume observed in *bed rest* and *weightlessness* of space pilots, is presumed to have a similar basis Procedures which decrease central blood volume, e g , *hemorrhage*, *positive pressure breathing*, have the opposite effect The effects on fluid volume are mediated by reflex effects on antidiuretic hormone and aldosterone release and on renal blood flow Right atrial receptors and arterial baroreceptors participate in these responses as well The importance of vagal receptors for these actions is clear since vagal cooling or vagotomy blocks many of them Thus, intrathoracic vascular receptors importantly represented in the left atrium, respond appropriately to central blood volume and pressure fluctuations, so as to regulate blood volume Figure 3 151 illustrates several of these effects in response to graded hemorrhage (see also p 3-235)

Other experiments in man have functionally implicated pressoreceptors in the thoracic portion of the central blood pool in adjustments of the circulation (Fig 3 152) When the legs of a recumbent man are passively raised, the blood flow through the skeletal muscles of the forearm is increased greatly as is the central pool of blood Since at this time, the arterial blood pressure remains the same or even decreases, this dilation in the forearm is evidently not a consequence of arterial baroreceptor stimulation Inflation of leg cuffs before trunk and leg elevation, or the use of sympathetic blockade to the forearm, eliminates the flow response While the precise nature of the stimulus and accurate location of the receptors concerned in the reflex have not been determined, it seems

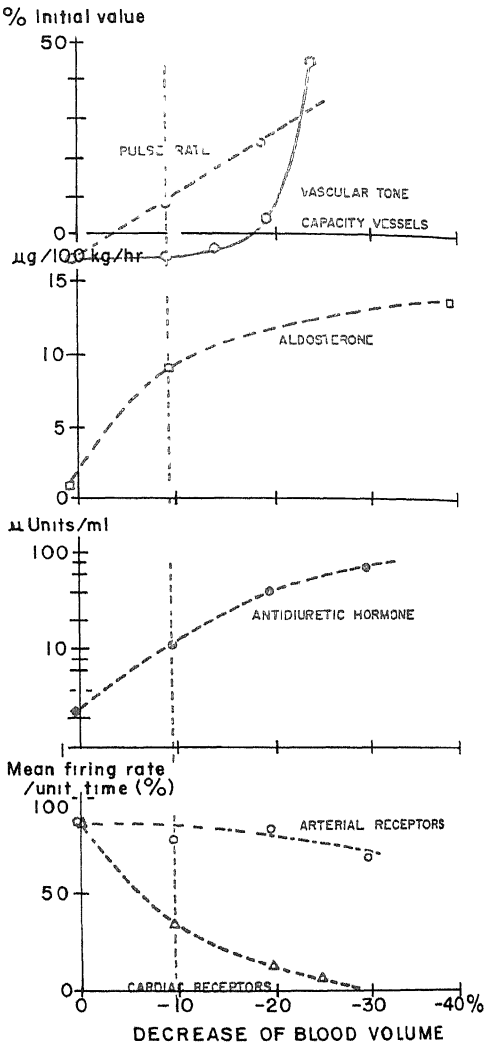


Fig 3 151 Responses to a graded hemorrhage With loss of the first 10% of blood volume the firing rate of atrial (cardiac) receptors falls to one-half As the loss exceeds 20% there is a fall in firing rate of aortic (arterial) baroreceptors The plasma aldosterone (μg released into the plasma) and antidiuretic hormone (μ units/ml) concentrations rise causing sodium and water retention The pulse rate increases steadily as blood volume falls There is little change in tone of capacity vessels until collapse threatens when it increases effectively to move blood to central regions (Redrawn from Gauer et al 1970)

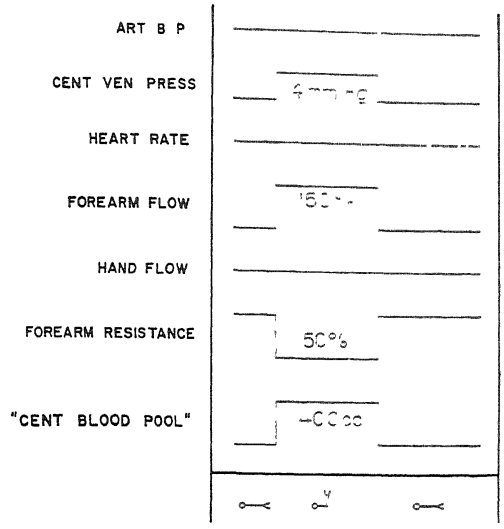


Fig 3 152 Passive elevation of the legs in recumbent man produces elevations in the central blood pool and central venous pressure with increased peripheral flow while heart rate and systemic arterial pressure are unchanged The decreased peripheral resistance is felt to be secondary to reflex dilation

likely that forearm dilation is due to stimulation of receptors in the low pressure area of the thoracic vascular bed Rapid positive and negative pressure breathing causes a similar augmentation of blood flow in forearm and calf without changes in arterial blood pressure and heart rate Presumably, this response is also to be explained through stimulation of pressoreceptors in the chest

Respiratory Reflexes. *Vasoconstriction in the Finger* Vasoconstriction occurs in a finger after voluntary deep inspiration or passive deep inflation of the chest with air under positive pressure The vasoconstriction was determined by a finger plethysmograph in which a blood flow decrease was noted with inspiration which bore no relation to the respiratory fluctuations of systemic blood pressure The receptors and afferent pathways of this reflex are unknown except that without doubt, they arise within the lungs

Respiratory sinus arrhythmia (increase in heart rate during inspiration, decrease during expiration) has been variously attributed to the influence on medullary cardioregulatory centers of impulses from· *1*) medullary respiratory centers, *2*) stretch receptors in the lung, or *3*) right heart volume receptors Experimentally, paralysis of respiratory movements and β-adrenergic blockade does not abolish the arrhythmia, while bilateral vagotomy does (Hamlin et al , 1966) Thus, sinus arrhythmia can result only from waxing and waning of vagal efferent activity in phase with periodic respiratory center discharge There is evidence that fluctuations in efferent sympathetic activity also play a role Periodic Hering-Breuer (q v) reflexes and right atrial distention during inspiration have not been ruled out as playing a contributory role In addition to the effect on heart rate, a positive inotropic action on left ventricular contraction during the acceleratory phase has been demonstrated

PERIPHERAL VASCULATURE REFLEXES

Although little is known concerning reflexes originating in the peripheral blood vessels, interest has been aroused in sensory mechanisms that may lie in the peripheral vasculature Haddy and Gilbert (1956), using anesthetized dogs in which pressures were measured in the brachial artery, cephalic vein, and a small artery and a small vein of the paw (with maintenance of a constant brachial artery flow by a pump), found that elevation of cephalic vein pressure produced dilation of the arteries and veins which was completely compensated for by constriction in the small vessel segment After procaine nerve block, the vessels respond like a passive elastic system Therefore, the observed small vessel constriction must have been of venous origin However, this local venovasomotor reflex has not been confirmed by Greenfield and Patterson (1956) using the plethysmograph

Summary

Change in activation of the vasoconstrictor fibers is responsible for blood pressure homeostasis and heat loss regulation of the skin blood flow Centrally induced vasodilation is caused almost entirely by inhibition of constrictor tone There may be a generalized or a strictly segmental or regional engagement, depending upon the type of stimulus Together with the sympathetic vasodilator fibers to skeletal muscles, vasoconstrictor fibers may affect specific vascular areas in a discharge pattern causing marked redistribution of blood flow The vasoconstrictor fibers are looked upon as the main neurogenic adjustors of the peripheral circula-

tion showing prompt and rational adjustments of their impulse discharge to any change in environment, especially those changes that affect the hypothalamus and the vasoconstrictor center in the medulla.

In the nerves from the pressure or stretch sensitive vascular walls of the aorta and carotid sinus, an increased number of impulses is discharged with each rise of blood pressure; the impulse frequency, systemic blood pressure, and heart rate can be varied at will when these areas are artificially perfused at different pressures and pulse pressures; section of the nerves results in a persistent hypertension. These presso- or baroreceptors are joined by the aortic and carotid sinus nerves with the nerve centers in the medulla oblongata, which maintain systemic blood pressure at normal levels. Any arterial blood pressure deviation causes compensatory reflexes, through the baroreceptors, to re-establish a normal blood pressure. The efferent pathways to the heart for these baroreceptor vasomotor reflexes are the vagus and the sympathetic; in addition, the sympathetic vasoconstrictor fibers adjust the size of the "resistance" and "capacity" vessels in control of peripheral resistance to blood flow and the return of blood to the heart. The sympathetic fibers which control secretion of epinephrine and norepinephrine by the adrenal glands also stimulate these secretions and add a blood-borne hormone to the nervous influence, but as already indicated, this humoral influence is almost insignificant compared to that of the direct nerve effect.

The baroreceptors also react to nonspecific pharmacological agents. Baroreceptors exist in several other tissues—common carotid artery, thoracic and mesenteric arteries; these are probably not related to blood pressure homeostasis, but may be engaged in local reflex adjustments of peripheral resistance and blood flow. Baroreceptor nerve endings are in the walls of the caval veins, heart (mainly the atria), and in the pulmonary circulation. Presumably, these receptors are primarily engaged in the reflex adjustment of venous return, heart rate, blood volume, and cardiac output. In the presence of marked hypotension from hemorrhage, chemoreceptors in the carotid and aortic bodies are excited to induce vasoconstriction in the resistance and capacitance vessels through the sympathetic constrictor fibers. Chemoreceptors are also present in the pulmonary and coronary circulations, but the physiological stimuli adequate for their activation are not known.

9 / REGULATION OF SYSTEMIC AND PULMONARY CIRCULATION

SYSTEMIC CIRCULATION

Arterial Blood Pressure

The first determination of arterial blood pressure was in 1733 when the Reverend Stephen Hales inserted a brass cannula into the central end of the femoral artery of a horse, connected it to a long vertical glass tube, and observed that a column of blood filled the tube to a height of 8 ft , 3 in (185 mm Hg)

ARTERIAL BLOOD PRESSURE IN MAN

Normal Values. The average systolic pressure of young male adults at mental and physical rest in the sitting position (as usually measured in a routine medical examination) is usually given as 120 mm Hg, the diastolic as 80, the mean pressure as 100, and the pulse pressure as 40 However, the range of normal blood pressures may be from 90 to 120 mm Hg systolic, and 60 to 80 mm Hg diastolic Slight diurnal variations in blood pressure of from 5 to 10 mm Hg systolic occur, the peak being in the afternoon and the lowest level in the early hours of the morning

Age, Sex, and Build. Age exerts a definite influence upon the blood pressure levels At birth, the systolic pressure measures from 20 to 60 mm with an average of 40 mm It rises rapidly, however, and has an average value of about 70 mm at the end of a fortnight and 80 mm at the end of a month A slow steady rise takes place from this time until about the 12th year when it averages 105 mm With the onset of puberty, a more sudden rise occurs, which in boys reaches 120 mm at about the age of 17 A steady though not great rise in blood pressure, from adolescence to old age, is the rule in health, the averages for the age of 60 being given as about 140 systolic pressure and 87 diastolic In women up to the time of the menopause, the systolic pressure is from 4 to 5 mm lower than for men of the same age At the menopause however there is a somewhat abrupt rise and the pressure remains a little above the male average from then on (Table 3 20)

Comparing groups of markedly obese and normal persons, the former have a rather pronounced increase of systolic pressure The incidence of abnormally high blood pressure (hypertension) is also definitely greater in overweight persons

The Effects of Digestion, Emotion, Exercise, and Posture The systolic pressure is influenced to a small but definite extent by meals A rise of from 6 to 8 mm is the usual effect, and this lasts for an hour or so There is little change in the diastolic pressure, if anything, it is reduced, presumably as a result of vasodilation in the digestive organs and skin

Quiet, restful sleep may be accompanied by a fall of from 15 to 30 mm in the systolic pressure The fall is most marked during the first hours, rising gradually again after this until the time of waking If the sleep is disturbed and accompanied by imaginary motor activities, there may be no depression of the pressure, but rather an elevation, in some instances to as high as 200 mm systolic and 105 mm diastolic Excitement, fear, worry, etc , markedly affect the arterial blood pressure, especially the systolic The effects are brought about through increased cardiac action and changes in the state of the vessels through impulses playing upon the cardiac and vasomotor centers in the medulla The liberation of adrenaline into the blood stream may also be a factor

Of all physiological conditions, exercise, if of a strenous nature, has the most powerful effect upon the arterial blood pressure During the muscular effort or even immediately before, i e , at the instant that the exertion is contemplated, the systolic pressure commences to rise and may reach a height of 180 or 200 mm Hg Except in well-trained persons or athletes, this

Table 3 20

The average variations of blood pressure (after Hunter's compilation of observations on a quarter million healthy Americans)

Age	Systolic Pressure	Diastolic Pressure	Pulse Pressure
10	103	70	33
15	113	75	38
20	120	80	40
25	122	81	41
30	123	82	41
35	124	83	41
40	126	84	42
45	128	85	43
50	130	86	44
55	132	87	45
60	135	89	46

After Gager, 1930

is invariably associated with a large increase in heart rate to 150 or 180/min In the trained individual, the blood pressure rise may be quite mild The diastolic pressure shows a less pronounced rise (100 to 110) so that the pulse pressure is increased In light exercise the diastolic pressure may remain at the normal level while the systolic rises several millimeters Immediately after the exercise the pressure drops momentarily to normal or even slightly below It then mounts rapidly to its previous high level, from which it gradually declines again, and in a healthy person reaches the normal within from 1 to 4½ min The evanescent drop in pressure is explained as being due to the sudden relaxation of the abdominal muscles The blood is drained into the venous reservoirs These when deprived of their support (abdominal muscles) have their capacity increased and the blood flow into the right heart is temporarily curtailed It is not until an appreciable time has elapsed to enable the increased venous capacity to become filled again by blood pouring in from the recently active muscles, that an adequate flow into the right heart is restored

The diastolic pressure is somewhat higher in the standing than in the sitting position and lowest in recumbency This change is found to occur whether the postural change is brought about actively or passively and is evidently an overcompensation for the gravity effect The systolic pressure usually rises but to a less extent than the diastolic, so the pulse pressure is reduced Pressures, however, taken at as short an interval as 10 sec after the erect position has been assumed, show that the initial effect is a fall of from 6 to 22 mm Hg in the systolic pressure This is sufficient to stimulate the carotid sinus and aortic mechanisms and cause increased vascular tone with a consequent compensatory rise in pressure Compensation is usually complete within 30 sec Reverting from the standing to the sitting or recumbent position has the reverse effect, fall in diastolic pressure and rise in pulse pressure In persons with an abnormally and habitually low blood pressure, the systolic pressure may actually rise in the lying-down position and fall when the subject stands The diastolic, on the other hand, is always lowered in recumbency and raised in the erect posture

The Systolic, Diastolic, Mean, and Pulse Pressures. Hales, in describing his experiment, speaks of the blood column after it had ceased to rise further in the tube, oscillating above and below a mean level To quote his own words, "When it (the blood) was at its full height, it would rise and fall at, and after, each pulse, 2, 3 or 4 inches" By recording with an adequate pressure manometric system, these fluctuations are seen as waves synchronous with the heart beats (see Fig 3 153) The crest of the wave represents the maximal pressure, corresponding to the contraction or systole of the ventricle, and is called the systolic pressure The trough of the wave, i e, the point of minimal pressure, coincides with the end of the

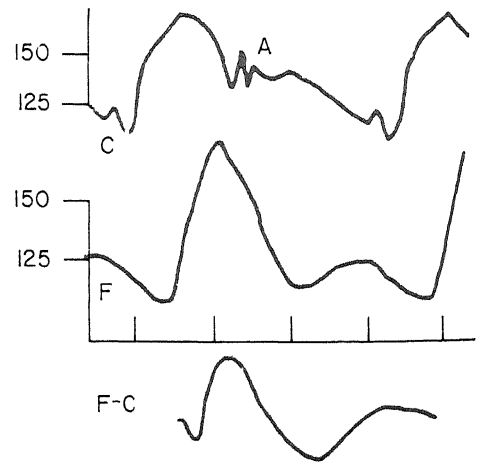


Fig 3 153 Arterial pressure pulses recorded from the central aorta *C* and the femoral artery *F* in the anesthetized dog Curve *F-C* beneath pulse tracings obtained by coordirectographic subtraction of the aortic pulse from the femoral pulse This represents the reflected or standing wave in the distal portion of the arterial tree Time 0 1 sec ordinate pressures in millimeters of mercury (After Alexander 1949)

resting phase or diastole of the cardiac cycle, and is called the diastolic pressure The mean pressure is usually given as half the sum of the values for the systolic and diastolic pressures This would be strictly accurate only if the pulse wave inscribed a perfect triangle which, however, is not the case The average pressure throughout the cardiac cycle, i e , the true or geometric mean, is generally somewhat lower than this, lying nearer the diastolic than the systolic pressure This can be obtained by integration of the curve

The difference between the diastolic and systolic pressures is the pulse pressure This is caused by the ejection of blood into the aorta during systole Its magnitude, other things being equal, will vary with the quantity of blood ejected by the heart at each beat

The systolic pressure reflects mainly the distensibility characteristics of the arterial system as it receives blood from the left ventricle It would be expected to undergo wider variations under the stresses of every day life than the diastolic pressure The latter represents the constant load which the arterial walls are called upon to bear and the resistance which the ventricular contraction must overcome to throw open the aortic valves Details of arterial pulse pressure changes are given on p 3-195 In the distributing arteries there is some amplification of the pulse pressure toward the periphery, but the two pressures tend to become more nearly equal in the terminal arteries, the pulse pressure there being reduced Since the pulse pressure is the difference between the systolic and diastolic pressures, it may be reduced by an alteration in one or other of these A rise in the systolic or a fall in the diastolic, will cause the pulse pressure to increase, while a fall in the systolic or a rise in the diastolic will lower the pulse pressure If both systolic and diastolic pressures rise or fall to an equal extent, the pulse pressure remains unchanged The mean pressure will rise as a result of a rise in either the diastolic or systolic pressure, or of both together, and will fall when a reduction of either or of both of these pressures occurs

DETERMINANTS OF NORMAL ARTERIAL BLOOD PRESSURE

Several factors combine to maintain normal arterial blood pressure They are, *1*) the pumping action of the heart, *2*) the peripheral resistance, *3*) the quantity of blood in the arterial system, *4*) the viscosity of the blood, and *5*) the elasticity of the arterial walls

Pumping Action of the Heart. The means by which the cardiac contraction exerts its effect upon the arterial blood pressure is, obviously through the *quantity* of blood which it is capable of discharging into the aorta in a unit period of time, i e , upon the output of the heart per minute or stroke volume times heart rate When more blood is forced into the already filled arterial system, it cannot escape at once from the system in the same amount as it is thrown into the aorta so the arterial walls become stretched The pressure rises until the velocity of flow through the arterioles is great enough to balance again the outflow from the system with the inflow Hales grasped this fundamental fact when he wrote, "the real force (pressure) of the blood in the arteries depends on the proportion which the quantity of blood thrown out of the left ventricle in a given time bears to the quantity which can pass through the capillary arteries (arterioles) into the veins at that time "

Peripheral Resistance. The peripheral resistance depends upon the caliber of the small vessels, mainly of the arterioles and, to a less extent, of capillaries, and upon viscosity of the blood It has also been demonstrated by Haddy et al (1957) that the macroscopic vessels (smaller arteries and veins) contribute significantly to resistance across systemic vascular beds By far, the greater part of the peripheral resistance of the circulatory system is constituted by the minute vessels of the muscles and of the abdominal structures In this connection the importance of the latter—the so-called splanchnic area—can be demonstrated by tying off all branches of the aortic arch, except the carotids, and the abdominal aorta below the inferior mesenteric branch, when little change in peripheral resistance results, on the contrary, an increase in the latter at once occurs when the blood supply to the splanchnic area is reduced by ligating the superior mesenteric artery Stimulation of the great splanchnic nerve, which innervates the rings of muscle fibers in the walls of these vessels, causes their constriction, and consequently a reduction in the outflow from the arterial system The pressure will continue to rise until inflow and outflow are again balanced. In the absence of compensatory changes in the other factors concerned in the maintenance of the pressure, this remains at the higher level so long as the constriction per-

sists Dilation of the vessels, i e , reduction in peripheral resistance, will of course be followed by the opposite effect When the vessels of the splanchnic area are fully dilated they are capable of accommodating almost all the blood in the body, in such an event the blood pressure would fall to zero The peripheral resistance might be compared to a dam in a river If the dam is raised or lowered, and no change occurs in the supply of water flowing down the river to the dam, the water continues to rise or fall, respectively (and its pressure in consequence increases or diminishes), until it reaches the new level From then on the quantity of water which overflows in a given time is the same as it was at the original level

The total peripheral resistance in animals or in man can be calculated from the mean blood pressure gradient (M) and the cardiac output (F), since all blood entering the aorta must, of course, pass through the peripheral vessels This can be derived for the whole body from the simultaneously measured arterial pressure and cardiac output, and for an organ or region from the local arterial pressure and flow There is a curvilinear relationship between pressure and flow in the aorta (Fig 3 154) This is also true in separate vascular beds where pressure and flow have been measured but the shape of the curves may be quite different

Various terminologies have been used to define peripheral resistance A *peripheral resistance unit* (PRU) has been defined as the flow in ml /sec /1 0 mm Hg pressure gradient

$$\text{PRU} = \frac{\text{mm Hg}}{\text{ml /sec}}$$

For a resting human with a cardiac output of 6 liters/min and mean blood pressure gradient from systemic arteries to veins of 100 mm Hg, the resistance would be

$$\text{PRU} = \frac{100 \text{ mm Hg}}{100 \text{ ml /sec}} = 1\ 0$$

Values range from 0 25 PRU with marked vasodilation to 4 0 PRU with intense vasoconstriction in man

In the pulmonary circuit with a mean pulmonary arterial pressure of 14 mm Hg and mean left atrial pressure of 4 mm Hg the gradient would be 10 mm Hg At the same cardiac output of 6 liters/min , the total peripheral resistance would be

$$\text{PRU} = \frac{10 \text{ mm Hg}}{100 \text{ ml /sec}} = 0\ 1$$

Values range from 0 03 PRU with exercise to 1 0 PRU in pulmonary hypertension

Many authors take such a ratio and by expressing M in dynes per square centimeter and F as milliliters per second, the peripheral resistance can be given in absolute units of force Thus,

$$\text{Peripheral resistance } (R) = \frac{M(\text{mm Hg}) \times 1332}{F(\text{ml /sec})}$$

$$R = \frac{\text{dynes/cm}^2}{\text{cm}^3\text{/sec}}$$

$$R = \frac{\text{dyne sec}}{\text{cm}^5}$$

1332 is a figure for the conversion of pressure to dynes

The peripheral resistance so calculated is found to vary inversely with the size of the animal, i e , directly with the surface area It amounts normally to from about 600 to 2000 absolute units in man, but may be over 5000 in arterial hypertension in which F shows little change The values for the dog and rabbit range from 2000 to 9000 and from 11,500 to

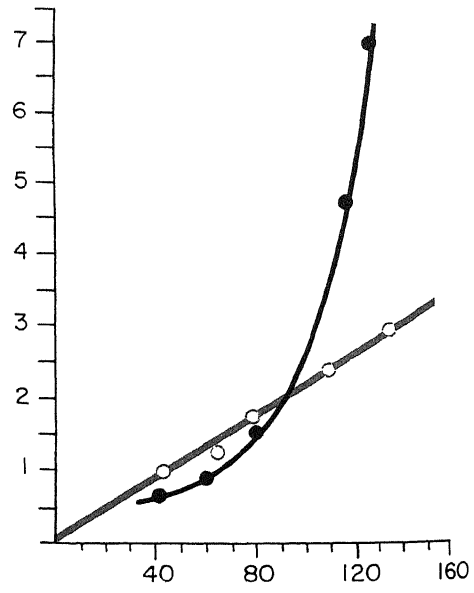


Fig 3 154 Relationship between mean arterial pressure in millimeters of mercury (abscissa) and cardiac output in liters per minute (ordinate) in a typical experiment Solid circles represent values observed with buffer nerves intact open circles values obtained after bilateral vagotomy and while a constant pressure of 210 mm Hg was exerted upon the isolated innervated carotid sinuses (From Levy et al 1954)

12,000, respectively This means that the minute vessels in a large animal offer less resistance than do the fewer number in a smaller animal, even though the vessels of both are constricted to the same degree This is due to the fact that the rate of flow in the smaller animal is greater in relation to the size of its vascular bed (cardiac output bearing a constant relationship to surface area which varies inversely with body weight)

Quantity of Blood in the Arterial System In any closed system of rigid tubes fluid must fill it to capacity in order that a pressure can be developed within it The arterial walls are distensible and elastic, and a certain degree of stretching of these must occur before any considerable pressure is created The arterial system must be actually over-filled, and the greater the extent of the over-filling, the greater will be the blood pressure Loss of blood, either of all its elements, as in hemorrhage, or of the fluid portion alone, if not compensated for sufficiently by readjustment of the other factors concerned in blood pressure maintenance, must inevitably result in a fall of pressure Increasing the total amount of circulating fluid artificially, as by the transfusion of blood or blood substitute, will elevate the pressure again

Viscosity of the Blood The greater the viscosity or "thickness" of any liquid the greater is the pressure required to force it along a length of narrow tube in a given time, or if the pressure remains constant the longer will be the time required for the liquid to traverse the tube The frictional resistance which is developed between the parts of the liquid itself, that is, the internal friction is greater when the viscosity is high than when it is low Viscosity depends upon the degree to which the molecules or particles of a fluid cohere Blood is some five times more viscous than water [1] With regard to the influence of viscosity upon the blood pressure, it is again a matter of outflow through the arterioles If the driving force remains constant and the caliber of the vessels is unchanged, then the greater the viscosity, the greater will be the frictional resistance developed in this region, and the less will be the quantity of fluid that will pass through in a unit of time The blood owes its viscosity to its colloids (plasma proteins) and to an even greater extent its suspended corpuscles, friction is developed between the surfaces of the latter and the surrounding fluid Changes in the concentration of the blood as a result of changes in its protein content or in the number of its corpuscles will therefore alter its viscosity

Elasticity of the Vessel Walls. This is concerned mainly with the origin and maintenance of the diastolic pressure and with sustaining the mean pressure at a higher level than would be possible in a rigid system under otherwise identical conditions The conversion of the pulsatile flow to a uniform one depends also upon the existence of a diastolic pressure

The physical principles involved in the maintenance of the diastolic pressure and the disappearance of the pulse beyond the arterioles may be best illustrated by a simple artificial model similar in principle to one devised by Borelli for the same purpose some 300 or more years ago In Figure 3 155 is represented a bulb syringe, *S*, valved at *V*, and having a short tube, *T*, which dips into a basin of water Leading from the opposite pole of the bulb is a longer tube When the bulb is alternately compressed and released, fluid is drawn from the basin and discharged from the mouth of the tube If the walls of the latter are composed of some rigid material (Fig 3 155*A*), it will be found that when the pump is worked the fluid issues from the tube in spurts or jets synchronous with each stroke, but no flow occurs between the strokes An increase in the frequency or force of the strokes does not alter the intermittent character of the flow nor does lengthening the tubing If the peripheral resistance of the vascular system be imitated by attaching a nipple of small bore

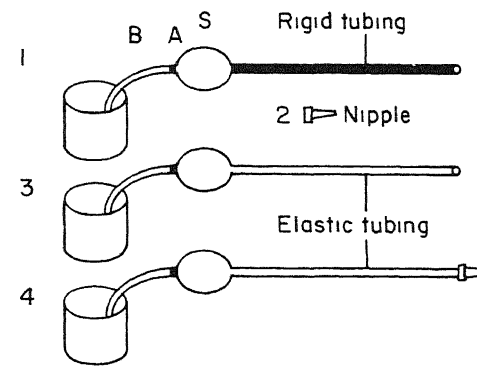


Fig 3 155 Demonstration of pulsatile and continuous flow

[1] This is an average figure Values obtained by different observers for the viscosity of blood, taking distilled water as unity, vary considerably but the majority range between 4 5 and 5 5

to the mouth of the delivery tube so as to increase the resistance to the outflow of fluid, the issuing stream is finer and its velocity is increased, but it still remains intermittent Let the elasticity of the arterial wall now be imitated by replacing the rigid tube by one of rubber, yet let the mouth of the tube be left free and not constricted in any way (Fig 3 155*B*) The intermittent character of the stream is unaffected However, if the small-bored nipple representing the peripheral resistance be fixed into the mouth of the elastic tubing the stream will be found to have lost its pulsatile character and to have become continuous (Fig 3 155*C*) Two factors are therefore necessary to produce this result, *a*) *resistance to the outflow*, and *b*) *elastic tubing* The reasons for this are clear If the fluid has free egress from the tube, most of that which enters it from the pump is discharged from the open end before the next beat occurs, the pressure, in consequence, does not rise to a sufficient height to distend the rubber wall, i e , elasticity is not called into play, overfilling of the tube does not occur, and in consequence the latter acts simply as though it were composed of rigid material

The foregoing facts apply directly to the arterial system The elasticity of the vascular walls and the peripheral resistance are both essential for the maintenance of the diastolic pressure As the contents of the ventricle are thrown into the already over-filled system during systole the added pressure which is then exerted upon the vascular walls causes their further distension After the completion of systole, the elastic walls rebound and, pressing upon the blood within their embrace, force it onward through the peripheral vessels In other words, the arterial lumen returns to its previous diameter and the energy that had been stored up during the stretching of the elastic tissue is in this way gradually expended during diastole

The elastic recoil of the arterial wall thus acts in a sense as a subsidiary pump to drive the blood onward in a continuous stream between the heart beats Otherwise, the pressure would fall to zero after each systole

It is clear then that any increase in the elasticity of the arteries, other factors remaining unchanged, will tend toward a lowering of the diastolic pressure If the aorta and its larger branches are stiffened (as a result of sclerotic changes), they cannot expand to the same degree as can healthy, resilient vessels, and therefore do not so readily accommodate the blood (60 to 100 ml) ejected from the heart during systole Such a state will lead to a rise in the systolic pressure Normally, however, the cross sectional area of the aorta increases considerably with age, so that it needs to expand much less to accommodate the systolic discharge

CRITICAL CLOSING PRESSURE

The finite intravascular pressure at which a vessel will close has been termed critical closing pressure It is measured by progressively lowering perfusion pressure and determining the level at which flow ceases This has been considered an index of vasomotor tone, although the subject remains controversial There is ample experimental proof that perfused peripheral vascular beds behave as though there exists a finite level of pressure at which resistance vessels simultaneously collapse, since flow ceases before driving pressure is reduced to zero Controversy exists over whether this mechanism plays a role physiologically In some situations, e g , the isolated rabbit ear perfused at low vascular tone, critical closure can be shown not to occur As L E Bayliss has suggested (Daly and Hebb, 1966), large numbers of resistance vessels often lie in parallel in tissues, and it seems unlikely they would all have identical critical closing pressures Thus, they would tend to close one after another and resistance should rise gradually, rather than becoming suddenly infinite, just as if the vessels were becoming gradually smaller In any case, the concept is quite generally used to explain cessation of flow through a perfused tissue when this occurs at perfusion pressures greater than zero

The idea of critical closing pressure has been evoked to explain, for example, renal failure in hypotensive shock Renal arteries are capable of high vasomotor tone which would increase reflexly with drop in arterial pressure, thus accounting for shutdown of kidney function when mean arterial pressures fall below 60 mm Hg or so

THE ARTERIAL PRESSURE PULSE

Velocity. The pressure change created by the ejection of blood into the aorta from the left ventricle is propagated throughout the arterial tree with a velocity considerably greater than the forward movement of the blood itself Were the vascular walls rigid, the impact of ventric-

ular ejection would be transmitted almost instantly, since liquids are incompressible However, the arterial walls are elastic and transmission velocity of the pulse wave is determined largely by vascular distensibility (i e , with the change in volume for a given change in pressure, $\Delta V/\Delta P$) The relationship may be expressed as follows $v = K(V\Delta P/\Delta V)^{1/2}$ where v is pulse wave velocity, V is initial volume, ΔV is increment in vessel volume, ΔP is increment in pressure and K is a constant (Bramwell and Hill, 1922) As vessel wall stiffness increases (decreased compliance), ΔV decreases and pulse wave velocity increases With increased blood pressure the arterial walls are more strongly stretched and pulse wave velocity increases (Fig 3 156) Similarly, the velocity is increased in old age when arterial walls are stiffened by arteriosclerosis The velocity in the young adult aorta is 3 to 4 m /sec and increases to 10 m /sec in the elderly In the stiffer peripheral arteries pulse wave velocity is 7 to 14 m /sec

Contour. The shape of the pulse curve changes as it travels peripherally (Fig 3 153) The high frequency components, such as the incisura, are damped out and disappear The systolic portion of the pressure wave becomes more narrow and its peak exceeds that of the central curve (i e , systolic pressure is higher peripherally) The diastolic limit of the pressure

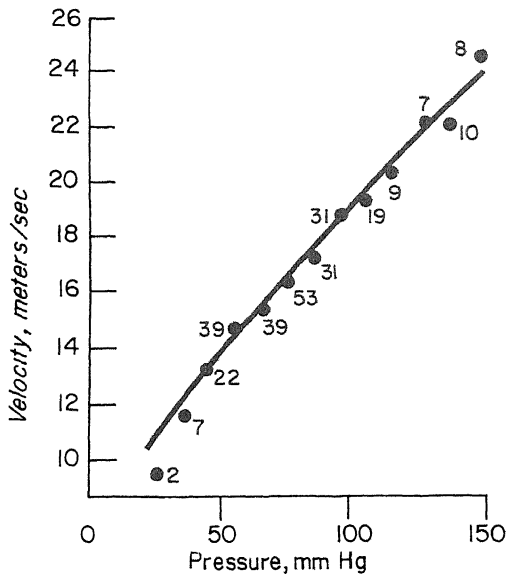


Fig 3 156 Curve showing the relation between arterial blood pressure (abscissa) and pulse wave velocity (ordinate) (From Landowne 1958)

pulse becomes lower, so that a mean pressure gradient accounting for blood flow prevails A hump appears in the diastolic part of the pulse wave The pressure wave passing along the arterial tree rebounds from areas of greater resistance, and travels retrogradely The magnitude of such a reflected wave can be roughly determined by subtraction of the peripheral pulse pattern from the central aortic pattern (Fig 3 153) Different parts of the pulse wave travel at different velocities, thus, contributing to the alteration in pattern from aorta to periphery The high pressure portion travels faster than the low pressure portion because arterial distensibility is less at higher pressure The result is crowding together of high pressure components at the periphery and peaking of the pressure pulse at the periphery (Fig 3 153*F*)

Determinants of Aortic Pulse Pressure Pattern. The aorta and its larger branches can be regarded as a large elastically distensible compression chamber with cardiac ejection entering at one end and drainage occurring at the other end The chamber possesses varying degrees of distensibility, conforming to those existing in various portions of the aorta and its branches The generated pulse pressure is caused by an imbalance between the ventricular outflow entering the aorta, and the peripheral run-off leaving the aorta with each systole The systolic portion largely expresses input into the aorta, the diastolic portion expresses decompression of the aorta or peripheral run-off, since, at this time, it is disconnected functionally from the ventricle Four factors are of major importance in determining the pressure and increments of pressure from moment to moment the elasticity or distensibility of the vessel walls, the volume of contained blood, the volume increment from cardiac ejection, and the decrement in volume caused by peripheral drainage These relations are expressed by the equation

$$E = (\Delta P/\Delta v)\ V$$

in which E = volume-elasticity coefficient, ΔP = change in pressure, Δv = change in volume, and V = capacity of aorta in diastole If the aorta and arteries constituted a simple elastic chamber, the pressure-volume relations (P/V) should be predictable from the preceding, and if P is known, V should become available It would be helpful if such a relation could be used to indicate what is happening to the blood

and whether the effect is centrally or peripherally induced However, difficulties in such an approach are *1*) there is no way to quantitate *V*, *2*) the pressure-volume relation of any part of the aorta is linear over only a limited pressure range, *3*) distensibility is different in various regions of the arterial tree, and *4*) changes have been revealed in the distensibility characteristics of the aorta by reflex stimulation and adrenalin injection

Usefulness of Aortic or Central Artery Pulse Patterns. Despite the difficulties in elucidating the pressure volume relations in the aorta, pressure pulses from the large arteries, depicting fairly well the pattern of the central aortic pulse pressure, have been found useful in predicting hemodynamic changes Such interpretation and extrapolation is based on experiments in dogs in which changes in peripheral resistance and cardiac activity have been deliberately induced and their effects on the pressure pulses separately determined Figure 3 157 shows the human subclavian pulse obtained indirectly The duration of systole and diastole can be determined In the presence of a high peripheral resistance and in a normal vessel, the pulse shows a much steeper rise throughout systole In the presence of a low aortic resistance but with good vascular volume, the curve has an initial early spike followed by a lower systolic tip, a rapid decline and a greatly reduced slope during diastole When the stroke volume is low, as in hemorrhage, the systolic part is small, smooth, and rounded, and diastole is almost flat Decreased distensibility of the aorta is revealed by a rapid, sustained, and large elevation of the pulse, a high incisura and a gradual decline during diastole Distinctive pulse patterns caused by lesions of the aortic valves are also shown Attempts to use central arterial pressure pulses to estimate stroke volume have been numerous but none has been successful

Arterial Blood Flow

With contraction of the human ventricle, a stroke volume or systolic discharge of 60 to 70 ml is ejected into the distensible aorta within less than 0 25 sec Of this, at least 45 ml is pushed into the aorta within the first 0 1 sec Space for this volume in the aortic arch is made by distending it and by removing some of its blood The resulting pressure change travels as a wave throughout the length of the aorta and its branches to the capillaries (and sometimes through them), and expands the entire arterial tree radially and longitudinally This enlargement results in a forward movement of blood from aortic valves to capillaries before the end of ventricular systole

PHASIC AORTIC FLOW

The phasic flow in the ascending aorta of the dog is roughly triangular in shape Fig 3 158*A* Acceleration reaches a peak ejection rate early

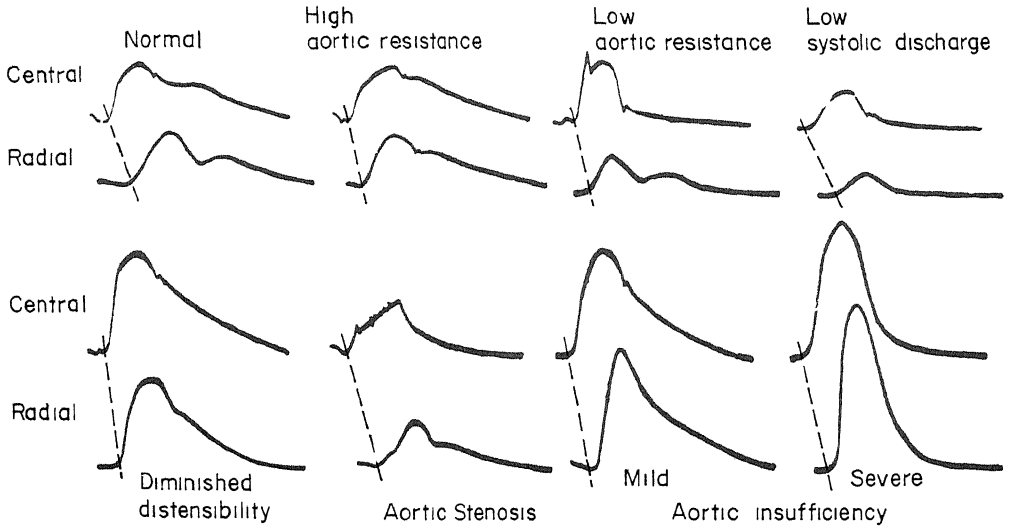


Fig 3 157 Illustration of changes in contour and amplitude of subclavian and radial pulses under abnormal cardiovascular conditions Discussion in text Note changes in transmission rates (Wiggers 1952)

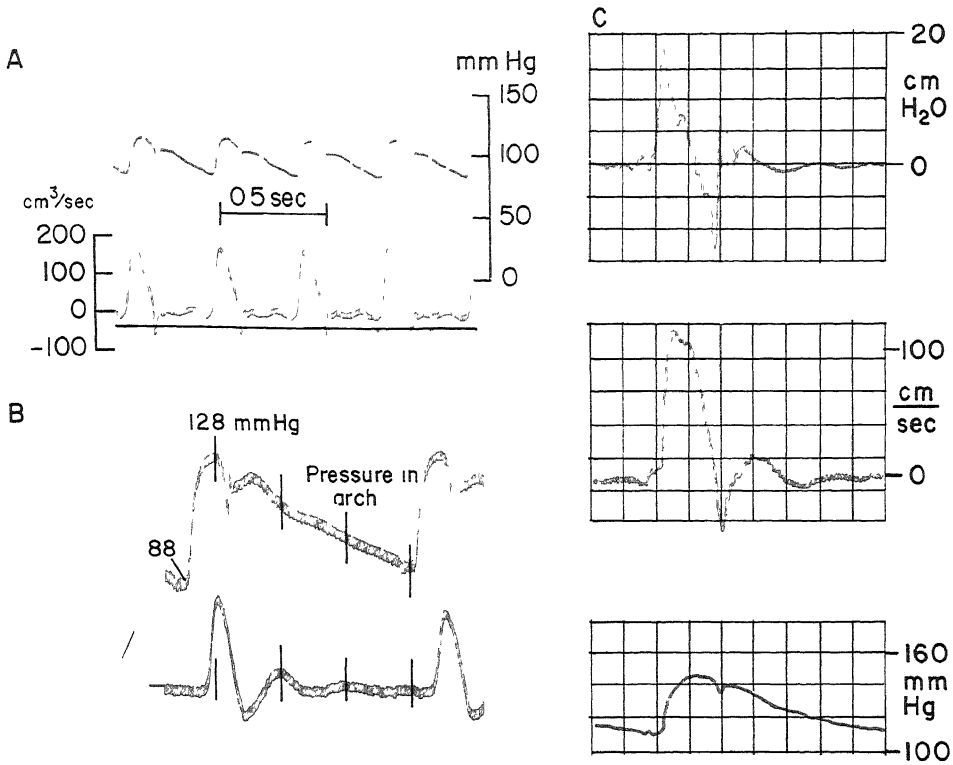


Fig 3 158 *A* Tracings from bottom to top flow in the ascending aorta pressure in the aorta (from Wetterer Minnesota Med 1954 37 79) *B* Tracing from top to bottom pressure in aortic arch flow in terminal aorta (from Spencer Circulation Res 1958 **6**, 492) *C* Simultaneous recording of *1*) aortic axial pressure drop between two points 6 cm apart at level of subclavian artery *2*) computed blood velocity at this level and *3*) aortic blood pressure at subclavian artery junction (From Fry et al 1957)

in systole following which there is a rapid deceleration ending in a momentary, small, high frequency backflow at the time of closure of the aortic valves The flow curve is flat with almost zero velocity during most of diastole The considerable modifying effect of the aorta upon the left ventricular ejection pulse is seen in Figure 3 158*B*, taken near the terminus of the aorta The flow pulse is now a series of smooth, highly damped oscillations of a frequency of 3 to 4 cycles/sec which continue to extinction, but which are renewed with each heart beat There is quite a large backflow in very early diastole Flow pulses in the isthmal area of the aorta (proximal portion of descending aorta) are intermediate in pattern between the central and abdominal patterns In Figure 3 158*C* are the calculated blood velocity curve and the aortic differential pressure curve, the latter taken by a double lumen catheter with pressure pick up holes 6 cm apart This shows the positive and negative acceleration waves which act on the aortic blood column to produce the flow pulsations The effects of inertia are seen throughout the cardiac cycle The beginning upstrokes of both pressure and flow curves coincide quite well, but when the first positive wave of the ΔP reaches its peak, flow acceleration is maximal, and when ΔP is maximally negative, flow deceleration is maximal This represents about a 90° phase lag

PHASIC FLOW IN PERIPHERAL ARTERIES

Flow patterns and simultaneous intra-arterial pressure curves characteristic of several peripheral arteries are illustrated in Figure 3 159 There are certain distinguishing characteristics A flow pattern is made up of waves whose directional changes have a qualitative correspondence with gradient changes in the simultaneously recorded intravascular pressure pulse and which may be characteristically distinctive for that artery and vascular bed Since the flow

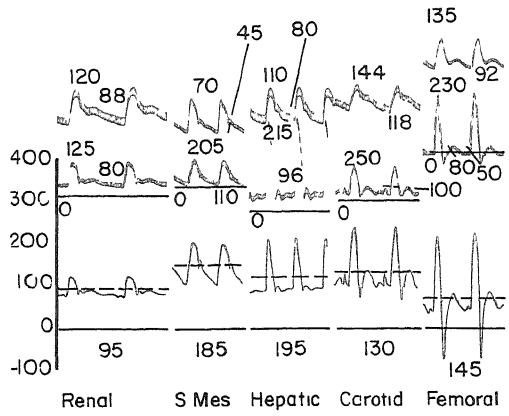


Fig 3 159 Phasic flow in peripheral arteries obtained by orifice meter (Shipley and associates 1943)

velocity varies with the differential pressure existing at the site of the flow meter, similarity in contour of the flow and applied pressure pulse is one criterion for the comparison of flow curves from different arteries Certain patterns, especially those of the superior mesenteric and renal arteries, have a well-rounded and sustained systolic portion in relative conformity to that of the pressure pulse Those of the hepatic and common carotid arteries are less well-rounded, those of the brachial and femoral arteries have a sharp systolic peak The flow pulse pressure is generally small in the renal and axillary patterns, and somewhat larger in the superior mesenteric and hepatic, while the femoral may be quite large Backflow components are consistently found in the femoral and axillary patterns and, on occasion, in the common carotid, while the renal, hepatic, and superior mesenteric flow curves exhibit only forward flow However, the main feature which permits a separation of the flow curves is the variable relation of the early diastolic to presystolic flow rates Comparison of this relation with similar early diastolic and presystolic points on the corresponding pressure curve, shows that the superior mesenteric, renal, hepatic, common carotid, and femoral patterns have, on this basis, a progressively graded dissimilarity to their respective pressure curves

ARTERY TO VEIN PRESSURE AND FLOW

In the quite small vessels (0 2 mm diameter) of the dog forepaw, while constant blood flow is maintained by a pump (Fig 3 160), it can be seen that with the environmental temperature at 20° C , the pressure drop from small artery to small vein approximates 40 mm Hg, and, from small vein to large vein, 6 mm Hg (Fig 3 161) Upon exposure to 0° C , the increase in pressure gradient is almost entirely through the small vessels with little gradient change between the small and large veins Since the blood flow was maintained constant, the increase in pressure gradient represents directly active vasomotor resistance changes

Venous Pressure

The superficial veins, which can be observed and whose pressure values and pulse pressure contours can be easily measured, serve as indicators of the functional status of the less accessible and deeper lying portions of the systemic venous system

THE VENOUS PRESSURE PULSE

The venous pressure pulse is best determined with the subject in the recumbent position Venous pressure is measured directly with an indwelling needle or catheter, or indirectly (without pressure values) by applying a cup over a vein external to the chest, such as the jugular, and connected to a manometer or a microphone receiver The venous pulse waves are usually timed against an electrocardiogram, arteriogram, or phonocardiogram (see chap 4) Two distinct types of pulsations, respiratory and cardiac, are seen in the peripheral and central veins With inspiration, blood is aspirated into the chest and the external veins tend to collapse while internal veins expand, with expiration, venous inflow is impeded and the external veins become distended while the chest veins tend to become smaller These

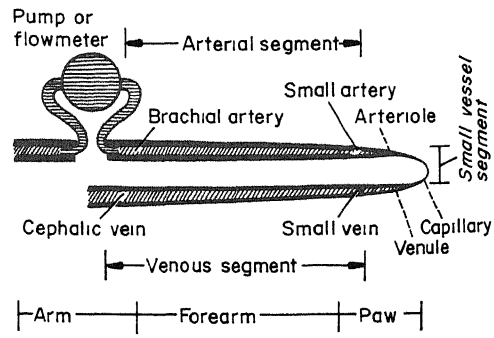


Fig 3 160 Schema showing means of measuring pressure gradient and resistance across small peripheral vessels (Haddy, 1958)

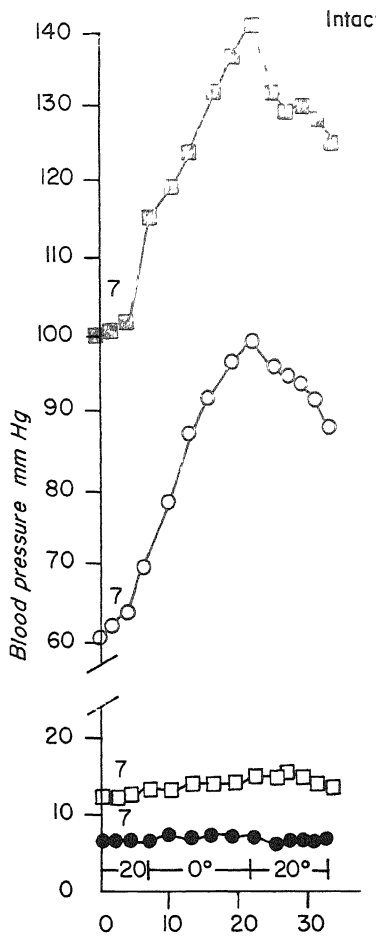


Fig 3 161 Effect of temperature change upon vascular pressures with flow rate constant in the foreleg of the dog *Ordinate* blood pressure *Abscissa* time in minutes and temperatues in ° C Numbers at each curve indicate number of animals in each series ● Cephalic vein □ small vein ○ small artery ■ brachial artery

changes make up the respiratory venous pulse The cardiac venous pressure pulse represents a retrograde reflection of a somewhat distorted version of the right atrial pulse pattern (see chap 4) These *a*, *c*, *v*, waves are clearly seen in the record obtained from a peripheral vein in Figure 3 162

NORMAL VENOUS BLOOD PRESSURE AND ITS VARIATIONS

Using the techniques previously described, mean blood pressure in the large, superficial veins of unanesthetized man at rest and in the supine position is found to be a function of distance from the right atrium The average value decreases from 13 8 mm Hg (188 mm H_2O) in the dorsal venous arch of the foot to about 5 2 mm Hg (70 mm H_2O) in the jugular vein The average pressure of 7 1 mm Hg (97 mm H_2O) in the antecubital vein drops to about 4 6 mm Hg (63 mm H_2O) in the vena cava opposite the right atrium (Fig 3 163) Thus, the overall venous pressure gradient approximates 9 2 mm Hg (125 mm H_2O) with a pressure difference of only a few mm H_2O occurring between the upper thoracic veins and the right atrium Although venous pressure measurements in the same subject are quite constant, considerable variation (up to 100 mm H_2O) may occur in the venous pressure from the same area in different persons

Peripheral and central venous pressures are altered in the same direction under many circumstances They do not appear to be related to age, sex, or hypertension, but they are decreased in severe hemorrhage and surgical shock They are increased by positive pressure breathing, the Valsalva maneuver, tilting, the excitement stage of anesthesia, the apneic

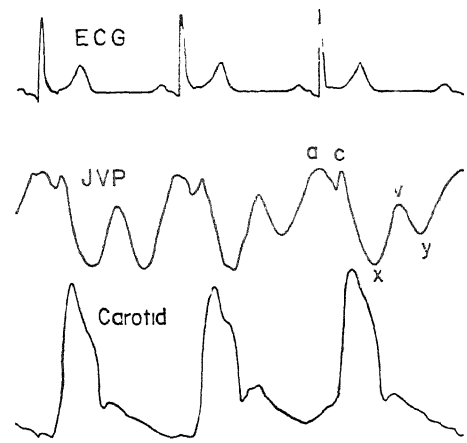


Fig 3 162 The jugular venous pulse pattern (*JVP*) of a normal subject An electrocardiogram (*ECG*) and carotid pulse tracing were simultaneously recorded The presystolic *a* wave is caused by atrial contraction The *c* peak is secondary to closure of the tricuspid valve although it may also represent transmission of the arterial pulse to the vein The descending limb of the pulse tracing after the *c* wave is secondary to emptying of the vein into the right atrium The negative *x* wave occurs during lowering of the atrioventricular septum and tricuspid valve during ventricular systole Then the pressure rises as the right atrium fills to form the *v* wave The second negative depression the *y* wave represents the fall in jugular venous pressure as the tricuspid valve opens and the right atrium empties (From Wood 1968)

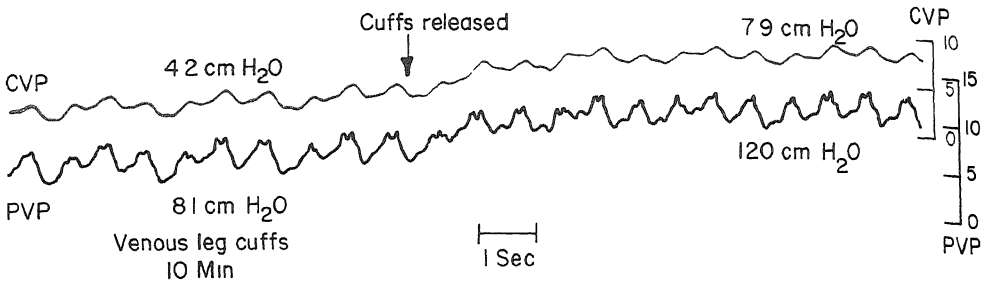


Fig 3 163 Comparison of central venous pressure (*CVP*) and the peripheral venous pressure (*PVP*) After release of occluding cuffs at the thighs the pressures return to control levels without a change in the pressure gradient (From Gauer et al 1956)

phase of asthmatic attacks, and in Cheynes-Stokes breathing The main pathological condition in which central and peripheral venous pressures are elevated, sometimes as high as 22 mm Hg (300 mm H_2O) in the antecubital vein, is congestive heart failure

DETERMINANTS OF VENOUS PRESSURE

The interplay of factors which determine systemic venous pressure and venous flow can be briefly stated The veins are placed between the capillaries or the AV anastomoses, and the right heart The blood enters the venules with an appreciable pressure residuum (vis a tergo) from the arterial side, after it has passed through the capillaries This pressure is met by the resistance to flow from the veins themselves and from the heart itself (the vis a fronte) The venous pressure is affected by internal factors which control the relation between the contained venous volume and the state of distensibility of its vessel walls Finally, there are various external or extramural factors (vis a latere), such as the actions of respiration, muscle, and gravity, which together represent the pressure effect of the surrounding tissues These factors act to insure that the pressure in the veins can never be less than the right atrial pressure, and whenever right atrial pressure increases, the peripheral venous pressure increases This excess of pressure (10 to 12 mm Hg or 125 mm H_2O) in the various veins, over that in the right atrium, is an expression of the pressure gradient (P_1–P_2 in Poiseuille's formula) This, together with a reasonable venous cross-section, causes blood to flow from vein to atrium

The Contraction of the Left Ventricle (the "Vis a Tergo"). *Vis a tergo* means "force from behind," and represents the pressure gradient from any point in the systemic circulation to the right atrium *Vis a fronte* means "force from the front," and has been used to represent any factor which affects the flow of blood into the right atrium Since both positive and negative pressure influences operate, usage of the term has become confusing It has been employed to designate *1*) the "back" pressure from the right atrium which impedes blood flow into the heart, *2*) the controversial suction effect of ventricular dilatation, *3*) the effect ("respiratory pump") of negative intrathoracic pressure which draws blood into intrathoracic veins, and *4*) the impediment to venous return imposed by positive intrathoracic pressure (e g , the Valsalva maneuver) (Guyton, 1963)

The energy of the ventricular contraction (the "vis a tergo") is expended in driving blood through the arterioles and onward through the capillaries and veins to the right atrium By the time the blood reaches the venules, the remaining pressure may approximate 7 to 8 mm Hg When it reaches the right atrium, the energy has been almost entirely dissipated in overcoming the frictional resistance offered by the vascular channels, at this point, the pressure may be less than 1 mm Hg

Quantity of Blood Flowing through the Arterioles in Relation to Capacity of the Veins. With dilated arterioles, the difference between the arterial pressure on the one hand and venous pressure on the other, tends to be reduced, with constricted vessels the pressure difference will be increased Rough estimates based on the volume-elasticity of the various compartments of the vascular system indicate that if a liter of blood is transfused into the circulation of man at rest, only 10 to 20 ml would be found in the arterial tree and about 980 ml in the extra-arterial system (systemic venous and pulmonary circuits), and that the pressure

elevation in the whole system would be only 4 or 5 mm Hg In bleeding and transfusion experiments in man, the circulation behaves like an elastic container, with pooling of blood in the extremities by occlusive cuffs, the central venous pressure falls without any apparent sign of counter-venomotor control This emphasizes the view that normal stresses to which the circulation is exposed, if they are not too abrupt or excessive, lead to only passive change in the resistance vessels and systemic venous system, and do not cause relaxation or strong contraction of the peripheral capacitance vessels

The veins are capable of adjusting their capacity to the quantity of blood received from the arterial system by reflex vasomotor activity The participation of venomotor tone in the control of the venous pressure and venous blood flow can be shown in animal experiments in which oxygen want, high CO_2, stimulation of the carotid sinus, and excessive hemorrhage are effective in reducing the size of the capacitance vessels In man, using strong stimuli such as the Valsalva maneuver, asphyxia, high G, or large blood losses, the veins constrict in what may be called an emergency reaction In the latter case, the central venous pressure may be normal or even increased

Central and peripheral venous pressures do not necessarily shift in the same direction The pressure gradient from very small veins (0 5 mm diameter or less) to large central veins, is not necessarily constant and can be markedly affected by various states With cold, hyperventilation, and moderate exercise, only the peripheral venous pressure may rise significantly This often independent action of local venous pressure and central venous pressure indicates that reflex venomotor activity of the capacitance vessels in normal persons is highly selective and regional in operation and that local small vein or venule pressure is not necessarily an index of happenings in the central venous system Hence, the state of distensibility of the arterioles, venules, and small veins, and large central veins, are not necessarily related, and net effects are unpredictable

The Action of the Right Side of the Heart upon the Blood Flow in the Veins. Obviously, if the blood is not passed on again by the right heart as quickly as it is carried to it by the great veins, the venous pressure will rise If the heart fails, the venous pressure rises and back pressure effects ensue

There is experimental proof that the activity of the right ventricle itself actually increases greatly vena caval flow into the right atrium This is illustrated in the vena caval flow curves of Figure 3 164, in which it can be seen that there is a larger forward flow of blood during right ventricular systole than during diastole During contraction, the descent of the atrioventricular junction enlarges the great venous reservoir which comprises the atria and venae cavae Obviously, this could be a very useful mechanism for maintaining venous return during tachycardia In slow hearts most of atrial inflow occurs during the long ventricular diastole, but during a rapid heart rate the proportion of atrial inflow due to active systolic "injection" of blood from the veins into the right atrium is greatly increased

The influence of cardiac action upon the movement of blood in veins and atria does not necessarily imply that the ventricles exert an aspirating action during diastole This problem has been debated many years and is considered on page 3-103 In general, it can be said that under strictly controlled experimental conditions, there is some positive evidence for this

Sometimes when vigorous respiratory efforts are made, slight fluctuations of the venous pressure can be detected in the peripheral veins of the human subject These variations rarely amount to more than 10 mm of H_2O, but may be considerably higher than this when dyspnea resulting from obstruction to the free entrance and egress of air from the lungs exists Owing to the inertia of the blood column, the aspirating effect is less evident, the nearer to the

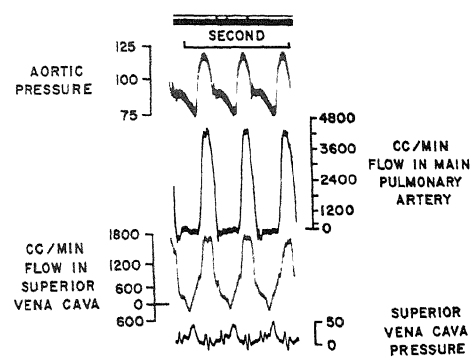


Fig 3 164 Simultaneous recording in open-chest dog of aortic pressure superior vena cava pressure and pulmonary artery pressure and superior vena cava flow with bristle flowmeter Note acceleration of venous return during systolic ejection (From Brecher 1956)

periphery at which the pressure measurements are made

If a forced expiration is made with the glottis closed (Valsalva's maneuver), the negative intrathoracic pressure can be abolished and a positive pressure of several millimeters of mercury substituted The veins of the neck, face, and limbs become distended with blood as a result of the impediment to the flow into the right atrium The peripheral venous pressure under these circumstances may rise to 400 mm H_2O or more In the converse maneuver of Mueller in which a forced inspiration is made with the glottis closed, the powerful suction effect may cause a fall of 50 mm H_2O or so in the venous pressure of a peripheral vein

Youmans et al (1963), have described an abdominal compression reaction in anesthetized dogs, thought to influence cardiac filling Tonic contraction of abdominal muscles is suspended during inspiration and restored upon its cessation Procedures which decrease blood volume in cardiac chambers lead to increased activity of the abdominal muscles without causing corresponding changes in amplitude of inspiration It is speculated that this reaction may serve to compress abdominal vessels and thus favor venous return

External Forces Natural and artificial respiration have a very large effect on central venous pressure and flow Direct measurements of flow with the bristle flowmeter in the closed-chest dog have shown flow to increase in both the superior and inferior vena cava during inspiration, and in turn, this effect is transmitted to the outlying veins (Fig 3 165) In the closed-chest dog, intermittent, positive-atmospheric-pressure artificial respiration (expiration occurring passively against atmospheric pressure) reduces significantly venous return to the right heart over that occurring with natural respiration Intermittent positive-negative pressure respiration (use of mild suction during expiration) augments vena caval flow over that occurring during positive-atmospheric pressure respiration In unanesthetized, normal humans, continuous negative pressure breathing increases cardiac output and stroke volume, while continuous positive pressure breathing has the opposite effect The untoward effects of artificial respiration arise when there is positive pressure applied during artificial inspiration, the intrapleural pressure, instead of decreasing, actually increases, and may even become positive This decreases vena caval flow secondary to an induced increase in pulmonary vascular resistance and, in part, to the mechanical compression or "tamponade" of the heart and lungs caused by the expanding lungs The mechanical aspects of respiration, however, are apparently not essential to maintenance of venous return or cardiac output, for in the presence of apneic oxygenation in dogs (a condition of induced ventilatory arrest with the trachea connected to a reservoir of oxygen), life continues and the cardiac output may even increase

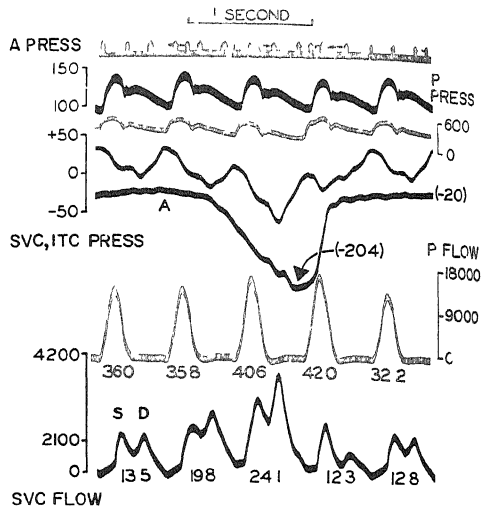


Fig 3 165 Effect of spontaneous respiration on venous return and cardiac output (closed-chest) Tracings from top to bottom time and base line aortic pressure in millimeters of mercury pulmonary artery superior vena caval (*SVC*) and intrathoracic pressure (*ITC*) in millimeters of water pulmonary arterial and superior vena caval flows in cubic centimeters per minute *A* beginning of inspiration *S* acceleration of superior vena caval flow during ventricular systole *D* acceleration of superior vena caval during ventricular diastole Stroke volume (in cubic centimeters) under pulmonary arterial flow curve Flow (in cubic centimeters) through superior vena cava during each cardiac cycle at bottom of record Electrical frequency response of both flowmeters reduced from 400 to 40 cycles/sec Superior vena caval pressure curve damped (Brecher 1956)

Posture and Exercise. The venous pressure varies, of course, as a result of the gravity effect, i e , with the position of the vein in relation to the right atrium, at which level, venous pressure is assumed to approximate zero (Fig 3 166) If man assumes the upright position and is motionless and relaxed, the measured pressure in an ankle vein approximates 85 to 90 mm Hg Similarly, the arterial blood pressure measured at the ankle will be increased over that at the heart level by an amount equal to

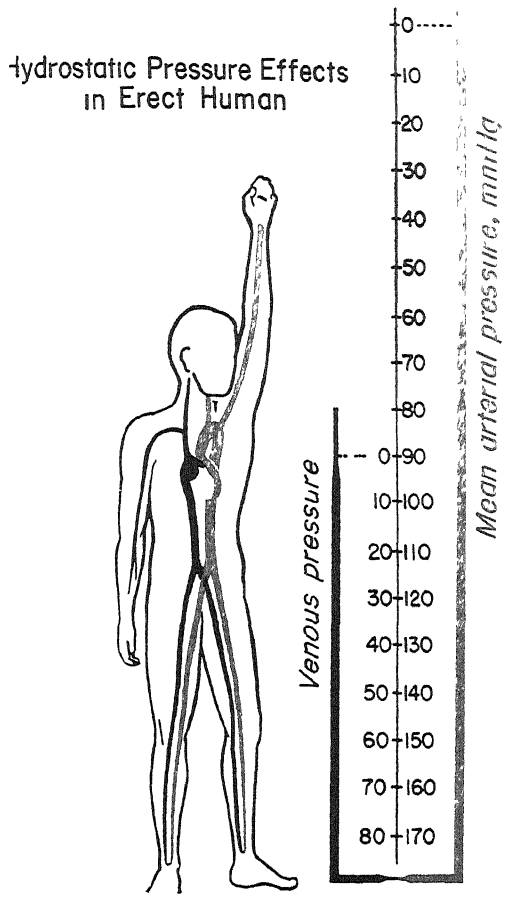


Fig 3 166 In erect position arterial and venous pressures are both increased by about 85 mm Hg at the ankle With the arm elevated over the head the arterial pressure at the wrist is about 40 mm Hg and the effective venous pressure is zero down to a level just above the heart (From Rushmer 1970)

the hydrostatic level Thus, the arterial and venous pressures at the ankle are increased by the same amount, and the pressure difference between vein and artery is the same at ankle and heart level This means that a favorable pressure gradient exists for venous flow Below heart level, the veins increase in fullness and the venous pressure gradient responsible for venous return from periphery to the heart is the measured venous pressure minus the hydrostatic column This is a small gradient but, since the veins are distended, resistance is quite low In the upright position, veins from the upper part of the body empty towards the heart These upper veins partially collapse and their intravascular pressure may approximate zero Above heart level, the pressure gradient along the venous channels is the height of the hydrostatic column above heart level This gradient is large but since the veins are partially collapsed, resistance is high

There is, however a special feature of the venous system which distinguishes it from the arterial system This is the tendency of the veins to collapse In the low pressure system, even in the recumbent position, the veins are not usually round but may be partially or fully, and intermittently collapsed For surface veins, collapse arises from increase in atmospheric pressure, in the deeper veins, it is caused by the surrounding tissue pressure acting on the venous walls This occurs during normal respiration in the neck veins, and in the inferior vena cava below the diaphragm where this process is aided by the increased intra-abdominal pressure With deep and prolonged inspiration, these veins may completely collapse Such a collapse tendency may be greatly augmented by an increased gravitational stress In the upright position, the pressure in the neck veins falls to 0 mm Hg, and the atmospheric pressure on the outside of the neck causes these veins to collapse all the way up to the skull This causes the pressure in these veins to remain at zero along their course It is this tendency for collapse that causes resistance in the large veins, for when distended, resistance is minimal However, not all veins are collapsible, especially the venous sinuses in the skull, vertebral column, bones, liver, and spleen, which are held open by firm-walled surrounding tissues For example, the veins in the skull are in a noncollapsible chamber, and in the standing position a negative venous pressure of about 10 mm Hg exists in the sagittal sinus because of the hydrostatic pressure difference between the top and base of the skull

In man, tilting into the feet-down position produces an immediate decrease of inferior vena caval flow which is more or less proportional to the degree of tilt The decreased venous return under such conditions is of short duration because of a number of compensatory vasomotor mechanisms which operate very quickly However, even with these compensations, the venous return is somewhat less than in the supine man In tilting the human subject into an upright posture, the blood pressure at the carotid sinus decreases mildly and the heart becomes smaller (Fig 3 167) About 400 ml of blood leaves the thoracic compartment to accu-

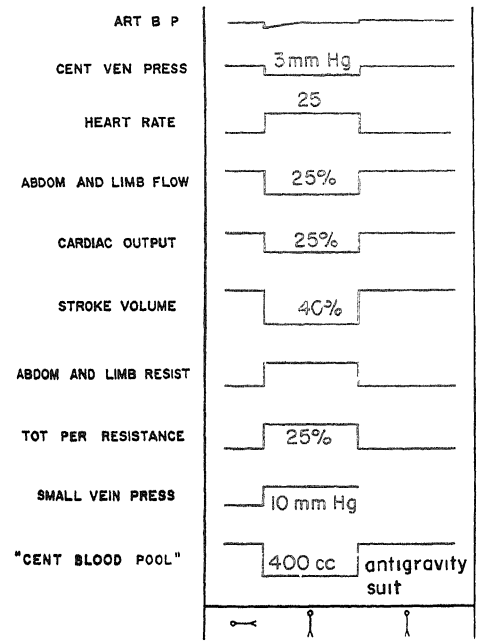


Fig 3 167 Effect on various cardiovascular parameters in man of changing from supine to the upright position

mulate mainly in the veins of the extremities (determined by leg plethysmographs) The pressure in them increases and the filling pressure of the heart drops Cardiac output drops mildly, stroke volume decreases considerably, the heart rate rises There is also a decrease in blood flow through skin, skeletal muscle, liver, and kidney These changes are at least partially explainable on a mechanical basis, i e , tipping into the erect posture might be regarded as a functional hemorrhage into the lower portions of the body Since the pressure in a superficial arm vein maintained at heart level rises considerably (about 10 mm Hg), peripheral venous constriction has taken place to shift blood centrally to make more blood available to the heart for maintaining cardiac output, although its effect is not enough to maintain central venous pressure That this limitation of peripheral blood flow and pooling are reflex phenomena is shown by its absence in a sympathectomized extremity The fact that application of a G suit before or after the subject has assumed the upright position, either prevents these changes, or re-establishes the original state of the circulation, mitigates against the role of the arterial baroreceptors and favors the stimulation of low pressure receptors While their precise stimulus and accurate location are not known, they presumably belong to the many receptors demonstrated anatomically in the thoracic vascular bed

The circulation of animals such as the domestic rabbit, eel, and snake, which have not acquired a compensatory mechanism, are placed at great disadvantage when the vertical position is assumed The eel and snake can be killed by immobilization in the vertical head-up position They literally bleed to death into their veins Thus, in the human subject if, upon assuming the erect posture after a protracted confinement to bed, the muscles of the abdomen and limbs are weak and the tone of the nervous mechanism governing the peripheral vessels lowered, the hydrostatic effect is overcome with difficulty The blood subsides into the capacious abdominal veins and capillaries, and the right heart is no longer adequately supplied with blood The arterial pressure falls and the cerebral blood flow becomes inadequate The subject turns pale, sweats freely, and feels giddy or perhaps nauseated, and in a complete faint loses consciousness (syncope)

In exercise, a much greater volume of blood enters the central veins from the arterial side The various mechanical assists already mentioned, that increase the venous pressure gradient and venous blood flow toward the heart and decrease the venous reservoir, now operate even more effectively In addition, the act of muscle shortening propels blood toward the heart

PULMONARY CIRCULATION

Regulation of Pulmonary Pressure and Flow

Uniquely interposed between the right and left heart and entirely enclosed within the negative pressure confines of the thorax, the pulmonary circulation exhibits certain pecularities uncommon to the other organ-perfused systems For instance, *1)* it is the only circuit in the body to receive the total cardiac output, *2)* it operates at a perfusion pressure one-seventh to one-eighth that of the systemic circuit, *3)* its mean transit time is equal to the systemic, albeit across a much reduced linear distance, reflecting the enormous surface area of the pulmonary capillary bed, *4)* it is predominantly controlled by the beat to beat pressure differential between the pulmonary artery (right ventri-

cle) and the left atrium upon which phasic, intrathoracic respiratory variations are superimposed, and 5) it is the only arterial circuit which carries desaturated hemoglobin and which, by virtue of the alveolar-capillary pressure difference, the desaturated hemoglobin and the huge surface area participates in large instantaneous gas exchanges Normal data for dog and man are given in Table 3 21

GENERAL CONSIDERATIONS

Bronchial Circulation. The blood supply to the bronchial connective tissue of the lung is part of the systemic circulation There is free communication between the capillaries of the pulmonary and bronchial systems, and these capillary beds may drain into either the systemic venous system through the azygos vein or through the pulmonary veins into the left atrium Thus, oxygenated pulmonary venous blood is diluted to a small extent (perhaps 1 to 2%) by unoxygenated bronchial venous blood The intermingling of the two circulations at the capillary level provides a potential shunt, which can serve to prevent elevation of capillary hydrostatic pressure should increase in either right or left atrial pressure occur unilaterally Under such circumstances, capillary blood can drain through the venous system with the lower pressure

The bronchial vessels can also provide collateral circulation to the lungs when the pulmonary arterial supply is inadequate Thus, in pulmonary artery atresia the bronchial arteries enlarge anastomoses with pulmonary arteries develop and poorly oxygenated blood reaches alveolar capillaries from the aorta

Pulmonary Resistance. To determine the mechanisms controlling pulmonary flow, it is necessary to know the resistance to flow Classically, this should be simply the ratio of the difference of pressure across any portion of a vascular bed to the blood flow that develops, it is a convenient factor by which the pressure drop can be divided to give the flow Such resistance gives information regarding the two factors upon which it depends, the viscosity and geometry of the vascular bed Since, however, the relation of pressure and flow is not a straight line and when plotted does not pass through the origin, the mechanisms involved are not determinable from a single point relationship To make the results interpretable, a vascular bed must be perfused in the control state and in the presence of a variable at the same pressure while flow is measured, or at the same flow while pressure is measured In no other way can it be even roughly determined that a calculated change in resistance (pressure divided by flow) reflects vasomotor activity produced by the stimulus, or reflects only the passive change in resistance caused by altera-

Table 3 21
Representative pulmonary hemodynamic values under basal conditions

	Dog (10 kg , Unanesthetized)		Man (75 kg , 1 85 m ^{2}, 40 Years of Age)	
	Mean	Systolic/ diastolic	Mean	Systolic/ diastolic
Pulmonary blood flow (liters/min)	1 5		6 10	
Radius, pulmonary artery lumen (cm)	0 77	±0 07*	1 35	±0 08*
Average velocity of flow, pulmonary artery (cm /sec)	13 4	47/0	17 8	85/0
Pressures (mm Hg)				
Right atrium	2		4	
Right ventricle		27/2		25/4
Pulmonary artery	14	27/9	15	25/10
Pulmonary capillaries	8		10	
Left atrium	4		7	
Pulmonary vascular resistance (dyne sec /cm ^{5})	534		105	
Pulmonary blood volume (ml)	110		440	
Pulmonary mean transit time (sec)	4 4		4 33	

From Milnor, 1968
* Maximal systolic/diastolic excursion around mean radius

tion in left atrial pressure or in pulmonary flow There is also an additional difficulty in measuring changing resistance in the pulmonary circulation, for here the relationship between pressure and flow is quantitatively quite different from that in the systemic circulation The difference is that with the same volume perfusing both circuits, the pressure drop from the pulmonary artery to the left atrium is only about one-fifth that from the aorta to the right atrium This is illustrated in Figure 3 168 where the differential pressure between the pulmonary artery and the pulmonary veins is only about 15 to 20 mm Hg This means that while the systemic resistance at the arterioles is high and these vessels have relatively thick muscle cuffs, the resistance vessels at the outlet of the small pulmonary artery branches have quite thin muscular coats with relatively large luminae The tone of the terminal pulmonary vessels is thereby relatively low so that a rise in pulmonary venous pressure is readily transmitted to the arterial side, or so that a rise in blood volume injected into the arterial side will increase greatly the volume in the lungs As a result of the high degree of distensibility of the pulmonary vascular bed, secondary to the opening up of fresh capillaries or further opening of those being used, the blood contained within the lungs shows wide variations under different conditions and on a passive basis

Pulmonary Artery "Wedge" Pressure. If a small catheter is advanced along the pulmonary arterial tree until it wedges in a small branch artery, blood flow behind the catheter is blocked, and fluid in the catheter perfuses the capillary bed beyond Then, the mean pressure in the catheter reflects that in the capillaries, and the synonym "*pulmonary capillary pressure*" has been applied to such "wedge" pressures Since the level of left atrial pressure influences that in the capillaries, arterial "wedge" pressure has been used as an indirect measure of the mean pressure in the left atrium However, large discrepancies between the two may prevail, so that pulmonary arterial "wedge" pressure may not give sufficiently accurate representation of the left atrial pressure for hemodynamic studies (Daly and Hebb, 1966)

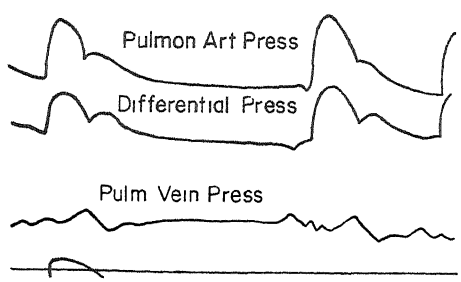


Fig 3 168 Record from the unanesthetized dog showing the pressure pulse contours in the pulmonary artery upper curve pulmonary vein lower and the differential pressure between the two middle curve Pulmonary artery pressure 28/10 mm Hg mean pulmonary vein pressure 10 mm Hg differential pressure 5 to 10 mm Hg (From Hamilton et al 1939)

Pressure Pulse in the Pulmonary Circuit. Pulmonary circuit pressures are much lower than those in the systemic arterial circuit In the unanesthetized dog the pulmonary artery pressure may be, for example, 40/10 mm Hg and the pulmonary vein pressure, 2 to 12 mm Hg (Fig 3 168) The pattern of the pulmonary arterial pressure is distinctive showing a marked incisura low on the dicrotic limb In late diastole, the pressure descent almost ceases and becomes horizontal

Pulse contours of the peripheral and central portions of the pulmonary artery display differences somewhat like those of the systemic circuit The pulse wave velocity in the pulmonary artery in man, studies during thoracic surgical procedures, is about 2 m /sec at normal pulmonary arterial pressure levels In patients with pulmonary hypertension, pulse velocity is increased to about 4 m /sec , indicating reduced distensibility of the pulmonary artery, since the pulse velocity is inversely related to distensibility The contour of the pulmonary vein pressure curve clearly shows a, c, and v waves, and closely resembles pressure pulses taken from the left atrium In open-chest dogs a 4 to 5 mm Hg pressure gradient was found between mean pulmonary venular (small pulmonary venous) and mean left atrial pressures

PULMONARY BLOOD FLOW

Phasic flow. In the pulmonary artery the flow of blood rises and falls more slowly than in the aorta (Fig 3 158A, 3.165) A small back flow occurs at the end of systole, as in the aorta (Fig 3 164) Flow remains pulsatile throughout arteries and capillaries (Fig 3 169), and small oscillations persist into the left atrium The dis-

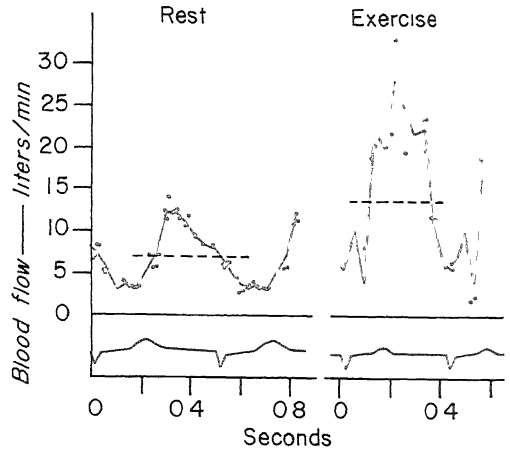


Fig 3 169 Simultaneous pulmonary capillary blood flow curves at rest and during exercise (DuBois and Marshall 1957)

tribution of flow within the lungs is affected by gravitational hydrostatic forces Thus, in the erect posture the lower lobes are perfused better The capillaries in the lung apices are about 15 cm above the heart base and are nearly or completely closed, those at the bases are distended and regional blood flow is greater here Thus, the normally low pulmonary artery pressure is barely adequate to perfuse the upper areas of the lungs The net effect of low capillary pressure combined with normal plasma oncotic pressure is to favor fluid reabsorption from the lung interstitial tissue Pulmonary edema appears first in the lower lobes, as would be expected because this is where hydrostatic pressure is highest

Sheet-Flow Concept. It has been suggested (Sobin et al, 1970) on morphological grounds, that the alveolar microcirculation is organized into a vascular sheet That is, an arrangement in which two walls of endothelium, held apart by connective tissue and cellular posts, enclose a sheetlike space, rather than the conventional model of a network of cylindrical tubes The interalveolar capillary bed would thus be visualized as an endothelium-lined, flat vascular sinus, with variable reservoir capacity through which blood advances as a moving sheet The validity of this concept as a better basis for understanding microcirculatory flow and gas exchanges across the capillary alveolar barrier has not yet been realized

Vascular Waterfall Concept. Some of the pressure-flow relationships in the pulmonary circulation suggest that under certain conditions arteriolar resistance alone is relevant, and the resistance of capillaries and veins is not In a waterfall, the flow is not influenced by the height of the fall, i e, the pressure drop, and vice versa, the pressure drop does not influence flow over the precipice In lung circulation perfused at constant pressure, flow maximizes when pulmonary venous pressure is slightly higher than alveolar pressure Changes in venous pressure much less than alveolar pressure fail to influence flow Since capillary pressure must exceed alveolar pressure for flow to occur at all through collapsible capillaries, lowering venous pressure must create a larger pressure drop between end-capillaries and pulmonary veins, while flow remains constant Thus a "waterfall" can exist between capillaries and veins (Permutt and Riley, 1963) Physiologically, this mechanism would operate only where arterial pressure is greater than and venous pressure less than alveolar pressure (Fig 3 170)

The influence of pulsatile arterial pressure on flow distribution in the lung appears to be frequency dependent (Maloney et al, 1968) Effective pulsatile pressure transmission to capillaries falls with increasing frequency of the arterial pressure pulse Thus, in the vertical lung, the height above the heart level at which blood flow will be detected decreases with increasing pulse frequency

In Figure 3 170, factors acting to influence distribution of blood flow in vertical lungs are diagrammed

CRITICAL CLOSING PRESSURE IN PULMONARY VESSELS

There is experimental evidence (Lopez-Muniz et al, 1968) that in the pulmonary circulation, critical closing pressure exceeding intra-alveolar pressure can be caused in both alveolar and extra-alveolar vessels by low pH or high CO_2 levels

EFFECTS OF RESPIRATION UPON PULMONARY AND SYSTEMIC BLOOD PRESSURES AND FLOWS

The effects of natural respiration on the hemodynamics of the pulmonary circulation are illustrated in Figure 3 165 The pulmonary arterial pressure falls during inspiration and rises during expiration One should expect that, as a result of the increased flow of blood into the right ventricle during inspiration (see superior

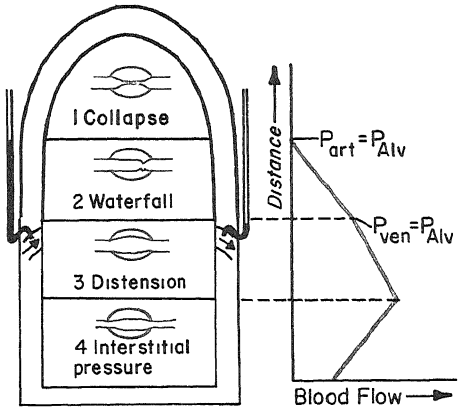


Fig 3 170 The resistance of vessels in different parts of the lung *1)* In this region the lung capillaries are collapsed if alveolar pressure exceeds arterial pressure and no flow occurs *2)* In this zone arterial pressure exceeds alveolar pressure and alveolar pressure is greater than venous pressure Flow is determined by the arterial - alveolar pressure difference (vascular waterfall) *3)* Both arterial and venous pressure exceed alveolar pressure and flow depends on arterial venous pressure difference Flow increases down this zone owing to vessel distension by gravity and to recruitment of vessels *4)* In the deepest zone flow decreases as the base is approached probably owing to interstitial pressure narrowing the diameter of extra-alveolar pulmonary vessels (Redrawn from Hughes et al 1968)

caval flow) and the greater systolic discharge, both the pulmonary pressure and pulmonary flow would rise during this phase of respiration However, only the pulmonary flow rises Because of the traction exerted upon the circumference of the pulmonary vessels by the surrounding lung tissue, their capacity is increased This more than compensates for the lack of change in the pressure gradient and for the greater amount of blood entering the pulmonary circuit during the inspiratory phase During expiration these effects are reversed The right systolic discharge is less but the capacity of the vascular bed of the lungs is at the same time reduced, an upward swing in pulmonary arterial pressure occurs With maximal expansion of the lungs, or during a forced expiration with the glottis closed (Valsalva's experiment), the vessels are strongly compressed by the surrounding lung tissue, and the pulmonary arterial pressure rises sharply

The increased capacity of the pulmonary vessels during inspiration reduces, momentarily, the flow of blood into the left atrium, the consequent reduction in the systolic discharge of the left ventricle causes a fall in aortic pressure (Fig 3 165) After a few beats of the right ventricle, the greater capacity of the pulmonary vessels again becomes filled and the flow of blood into the left chambers of the heart increases, the aortic pressure rises The succeeding expiration, by reducing the capacity of the pulmonary vessels, drives blood to the left side and further increases the discharge into the aorta, the systemic pressure, in consequence, continues its rise until near the end of the expiratory phase If the respiratory movements and the systemic blood pressure are recorded simultaneously, it is found that the blood pressure commences to fall at the commencement of inspiration and reaches its lowest point in the latter half of this phase, the blood pressure tracing then commences to rise and reaches its maximum toward the latter part of expiration In addition to these mechanical effects of respiration on systemic flow and pressure, irradiation of impulses to the vasomotor center from the respiratory center during inspiration can increase arterial pressure These are correctly called *Traube-Hering* waves after their discoverers (Schweitzer, 1945) Also, reflex pressoreceptor and chemoreceptor influences accompanying fluctuations in pressure and blood oxygenation with the respiratory cycle can contribute to the net effect

Distinguished from the respiratory waves in systemic arterial blood pressure are slower fluctuations called *Mayer* (sometimes erroneously referred to as Traube-Hering waves) waves, occurring at rates of from 1 to 3/min These represent oscillations in pressoreceptor and chemoreceptor reflex control systems, which are exaggerated under conditions of hypotension and tissue hypoxia Elevation of cerebrospinal fluid pressure elicits an increase in arterial blood pressure, the *medullary ischemic response of Cushing* (pp 3-179 and 3-211) When the blood pressure becomes high enough, ischemia is relieved and the pressure falls again until tissue hypoxia builds up once more, and the cycle is repeated Similar oscillations from cyclic carotid and aortic body chemoreceptor stimulation are seen especially in hypotension following hemorrhage The sinoaortic baroreceptors likewise participate in generating these waves, and may be the primary reflex influence under certain circumstances

EFFECTS OF EXERCISE ON THE PULMONARY CIRCULATION

The general form of the curve relating pulmonary artery pressure and pulmonary blood

flow in man and the intact dog at rest and exercise is shown in Figure 3 171 The curve indicates that pulmonary artery pressure is maintained below the upper limit of normal until the flow exceeds at least three times the base level, whereupon it increases progressively This absence of pressure rise with increased flow implies an expansion of the vascular bed from opening of new channels, widening of those already perfused, or a combination of both Thus, the resistance decreases It is believed (without proof) that this is on the basis of passive distension of the lung vasculature, especially since the same relation holds when attempts are made to denervate the lungs, when isolated lung lobes are perfused, or when variations in the quantity of blood in the lungs is induced by different procedures

As pointed out earlier, to understand what is happening hemodynamically in the pulmonary circulation, it is more important here than in any other vascular bed to know both the pulmonary artery pressure and pulmonary vein (or left atrial pressure) This is because the pressure gradient is only a few millimeters of mercury, and a 5 mm Hg-change in the pulmonary venous pressure could be transmitted back as a 5-mm Hg rise on the pulmonary artery side without any active change in peripheral resistance in the lung vessels So, unless it is known what is happening in the pulmonary veins, one

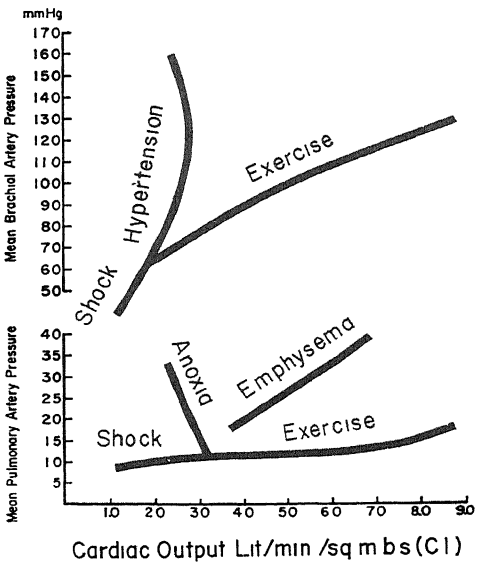


Fig 3 171 The relationship of pressure to flow in the systemic circulation and pulmonary circulation in normal and abnormal states Note the difference in response of the two circulations to exercise

is at a loss to ascribe the cause of a change in pulmonary artery pressure

Factors which passively affect pulmonary vascular pressure volume and flow include *1)* cardiac output, *2)* systemic arterial pressure, *3)* bronchial blood flow, *4)* ventilation, *5)* bronchomotor activity, and *6)* release of hormones (e g , catecholamines) into the circulation Active pulmonary vasomotion can be demonstrated by causing hypoxia, which produces vasoconstriction, or infusing acetylcholine, which causes vasodilatation and alterations in vascular tone are increasingly being recognized as important in the pulmonary circulation However, pulmonary blood volume changes are primarily passive in response to alterations in vascular distending pressure Thus, pulmonary blood volume increases in exercise, systemic vasoconstriction, impaired left ventricular function, and infusion of fluid It decreases with fall in systemic arterial pressure (vasodepressor reactions), improved left ventricular function, or decrease in total blood volume (hemorrhage)

Evidence for and against Vasomotor Control of the Pulmonary Circulation

The pressure in the pulmonary artery may increase as the result of back pressure from the left heart, an increase in pulmonary blood flow and vasoconstriction in the pulmonary vascular bed Also, pressure changes follow alteration of intrathoracic pressure

The pulmonary vasculature could respond to active vasoconstriction and contribute to an increase of pulmonary artery pressure The small pulmonary vessels possess muscular coats and are equipped with dual nerve supply—sympathetic and parasympathetic, stimulation of the pulmonary nerves increases the pulmonary vascular resistance, stimulation of the baroreceptors gives increased pulmonary flow and decreased pulmonary arterial pressure, casts of the two sides of a lung are markedly different if one side is perfused with noradrenaline up to the time of plastic injection (the latter showing marked constriction in the vessels of less than 25 μ in diameter Narrowing and gnarling of the small, muscular pulmonary arteries can be demonstrated angiographically in response to hypoxia and to the intrapulmonary arterial injection of catecholamines and serotonin in the closed-chest dog There is little doubt that wide variations in vascular resistance can normally occur in a purely passive way However, active changes in sympathetic

tone causing alterations in stiffness of large pulmonary arteries have been implicated in rapid circulatory adjustments (see pp 3-22 and 3-125)

In normal subjects, hypoxia, induced by breathing 12% oxygen in nitrogen, considerably increases pulmonary arterial pressure, mildly increases cardiac output, and is without effect on left atrial pressure or central blood volume This suggests that the vasoconstriction induced is on an active basis Although generalized, such vasoconstriction of the pulmonary bed would seem to serve no useful purpose, local hypoxia, however, could be helpful in regulating the distribution of blood by causing vasoconstriction which, in turn, diverts blood from the affected region to vessels in better aerated parts of the lung In experiments in man and animals in which one lung was respired with an hypoxic mixture, the lung probably developed vasoconstriction, since the arterial saturation returned to normal This would be the case if the vessels in the hypoxic lung constricted and forced the blood normally flowing through them into well-oxygenated channels Final proof, however, that hypoxia induces active vasoconstriction in man is contained in the experiments in which acetylcholine (a drug with strong vasodilating effects on the systemic circulation) was infused into the right atrium, but in such quantities that no systemic effects were produced In the presence of hypoxia induced by breathing a low oxygen mixture, infusion of acetylcholine decreased the elevated pulmonary arterial pressure but did not change the left atrial pressure, cardiac output, systemic blood pressure, or heart rate This seems to indicate that an active vasoconstriction caused by hypoxia is relieved by the drug This view is further substantiated by the observation that in emphysematous patients who represent a chronic and more limited form of hypoxia with some degree of arterial unsaturation, acetylcholine infusion lowers still further the arterial oxygen saturation The site and mechanism of action are unknown but they could be many—a local reflex from alveoli or pulmonary veins, chemoreceptor stimulation of the sinus or aortic body, direct local action on pulmonary bed, and shift of systemic blood into the pulmonary circuit, etc Presumably, the action is not related to sympathetic nervous control since the vasoconstrictor effect of hypoxia can be observed in man with a total sympathectomy Serotonin (5-hydroxytryptamine) infused into the intact dog produces a considerable rise in pulmonary arterial pressure with only a modest increase of cardiac output and no change in pulmonary venous or left atrial pressures, indicating active vasoconstriction

Summary

In summary, the pulmonary vessels are capable of considerable passive dilatation as part of their function as a blood reservoir Pulmonary blood volume and vascular resistance, thus are adjusted so as to minimize pressure changes as, for example, in exercise when blood flow can increase several fold with little increase in pulmonary arterial pressure Factors which increase systemic arterial pressure often have little effect on pulmonary arterial pressure (e g, pressor reflexes and pressor drugs), and pulmonary hypertension ordinarily does not accompany essential hypertension However, there is evidence suggesting that sympathetic tone variations alter vascular compliance in the lungs, thus affecting pulsatile flow and pressure transmission between the two ventricles In this sense, the pulmonary vascular bed probably serves as more than a passive conduit, but rather participates directly in maintaining equality of right and left ventricular output (see pp 3-124 and 3-21) The local vasoconstrictor effect of hypoxia serves to shunt blood away from poorly ventilated lung areas to those where oxygenation can take place In the vertical position, gravitational effects govern blood flow distribution in accordance with hydrostatic principles Thus, flow is minimal in the apices, increases at heart level and below until the most dependent areas are reached, where vascular resistance is elevated owing to increased interstitial fluid pressure (Fig 3 170)

CIRCULATION THROUGH BRAIN, SKIN, AND SKELETAL MUSCLE

Cerebral Circulation

Cerebral blood flow is determined by two opposing sets of forces the effective perfusion pressure and cerebral vascular resistance (see Fig 3 172) Effective perfusion pressure is the gradient between arterial blood pressure and internal jugular venous pressure, cerebral vascular resistance is defined here as the sum of all factors opposing the flow of blood through the brain, such as intracranial pressure, blood viscosity and vascular diameter

Internal jugular venous pressure for all practical purposes is negligible as compared to arterial pressure Even the obstruction of the vein by means of a tourniquet placed on the neck, producing pressures up to 300 mm of water, has no effect on cerebral blood flow (Moyer et al , 1954) Thus, arterial blood pressure is the sole important determinant of the driving force of blood flow across the brain

In man the cerebral blood flow may be determined by the nitrous oxide method (Kety and Schmidt, 1948) which is based on the Fick principle (see sect 3, chap 7) The subject breathes a gas mixture containing 15% N_2O, while blood samples are obtained from a peripheral artery and from the internal jugular vein (representing mixed cerebral venous blood) Cerebral blood flow per unit weight of brain may be calculated from the arterial and venous gas concentrations and the partition coefficient for nitrous oxide between brain and blood Results are usually expressed as milliliters of blood per 100 g of brain per minute, in resting, normal young men the mean value is 54 ml Assuming a weight of 1400 g for the brain, total cerebral blood flow is about 750 ml./ min

The relation of arterial blood pressure to cerebral blood flow is determined by the level of the blood pressure itself A lowering of the mean arterial pressure to a level of about 60 to 70 mm Hg by means of high spinal anesthesia, paralyzing most of the sympathetic vasoconstrictor nerves has no effect on cerebral blood flow which remains constant (Kleinerman et al , 1958) This indicates a compensatory lowering of vascular resistance in the brain (see below) Similarly, an elevation of blood pressure is without effect on blood flow due to a secondary increase in resistance (Moyer et al , 1954) This autoregulation of the cerebral circulation becomes ineffective, however, when mean blood pressure falls below 60 mm Hg In this hypotensive range, a lowering of blood pressure results in a reduction of blood flow Initially, this is compensated for by increased oxygen extraction by the tissues, but at mean pressure levels of about 30 mm Hg when cerebral blood flow is reduced to about 30 ml / 100 g of tissue/min , signs of cerebral ischemia appear (Finnerty et al , 1954) Conversely, raising the blood pressure under these conditions increases blood flow, unlike in the normotensive range

As seen in Figure 3 172, several extravascular factors contribute to "cerebral vascular resistance" Intracranial pressure, exerted upon the freely collapsible capillaries of the brain is a force opposing the flow of blood Any increase in this pressure, resulting from an expansion of the contents of the rigid cranial box (such as a brain tumor) elevates cerebral vascular resistance in a direct, linear manner (Kety et al , 1948) In spite of this, cerebral blood flow does not diminish until intracranial pressure reaches very high levels (about 450 mm of water) The maintenance of normal flow is due to an increase of systemic blood pressure brought about by peripheral vasoconstriction (see pp 3-208 and 3-179) It has been suggested that the stimulus for the latter is ischemia of the medullary vasomotor center

Viscosity of the blood increases resistance to flow in polycythemia cerebral blood flow is considerably reduced, it is restored to normal levels after correcting the increased viscosity by phlebotomies (Nelson and Fazekas, 1956).

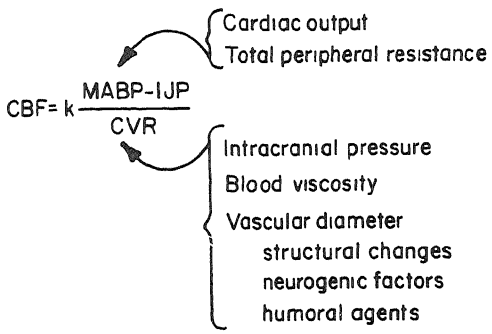


Fig 3 172 Factors determining the cerebral blood flow CBF cerebral blood flow MABP mean arterial blood pressure IJP internal jugular venous pressure CVR cerebral vascular resistance (From Kety 1958)

In anemia, on the other hand, cerebral blood flow is increased in proportion to the decreased red cell mass, transfusion produces a return to normal flow (Robin and Gardner, 1953)

The most important component of cerebral vascular resistance is the diameter of the cerebral blood vessels Their adjustment is responsible for the remarkable constancy of blood flow over a wide range of perfusion pressures, discussed above The exact mechanism of vasomotor regulation in the brain is not known Although the cerebral vessels are supplied by sympathetic fibers and norepinephrine is a powerful cerebral vasoconstrictor (King et al, 1952), there is no evidence to show that they are physiologically active in man Bilateral blocking of the stellate ganglia fails to produce any change in cerebral vascular resistance or blood flow, contrary to what might be expected if these nerves were normally contributing to the maintenance of cerebral vascular tone

RESPONSE TO HIGH CO_2

In contrast to this is the marked sensitivity exhibited by the cerebral vasculature to changes in arterial concentration of carbon dioxide and oxygen Carbon dioxide is the most powerful cerebral vasodilator known Inhalation of a gas mixture containing 5 or 7% CO_2, which raises arterial P_{CO_2} by 9 mm Hg, causes a 75% increase in cerebral blood flow Systemic blood pressure rises without any change in cardiac output, indicating a peripheral vasoconstrictor effect, but cerebral vascular resistance falls, showing the specific vasodilator action of carbon dioxide on the cerebral vessels (Kety and Schmidt, 1948) Hyperventilation, which lowers arterial P_{CO_2}, reduces cerebral blood flow

These effects of carbon dioxide are not related to associated changes in hydrogen ion concentration Intravenous infusion of sodium bicarbonate, which produces an increase of the total carbon dioxide content of the blood and a systemic alkalosis, has the same qualitative effect on cerebral vascular resistance as inhalation of 5% CO_2, which also raises total blood carbon dioxide but results in acidosis Conversely, an infusion of NH_4Cl duplicates the cerebral effects of hyperventilation both procedures reduce the total carbon dioxide content of the blood, but the accompanying pH changes are in the opposite directions (Schieve and Wilson, 1953)

The vasodilator effect of carbon dioxide is a threshold phenomenon No change in cerebral vascular resistance is detectable for arterial P_{CO_2} increases of up to 4 5 mm Hg, beyond this point there is progressive vasodilation with further arterial P_{CO_2} increments (Patterson et al, 1955) Decreasing the arterial CO_2 tension below normal produces cerebral vasconstriction, but the reduction in cerebral blood flow is proportionately less than the increase resulting from hypercapnia (Wasserman and Patterson, 1961)

In view of the pronounced cerebral vasomotor activity of carbon dioxide on the one hand, and the absence of any demonstrable neurogenic control on the other, it is logical to postulate that cerebral blood flow is regulated, at least in part, by local carbon dioxide concentrations It was calculated that a reduction in blood flow of about 30% would be required for capillary carbon dioxide tension to reach the average vasodilator threshold level of 4 5 In all likelihood, however, a reduction of this magnitude is not necessary, since under physiological conditions any increase in carbon dioxide concentration would be accompanied by hypoxia (see below)

RESPONSE TO LOW O_2

The cerebral vasomotor action of oxygen is qualitatively the direct opposite of that produced by carbon dioxide Inhalation of a gas mixture containing 10% oxygen, which reduces arterial oxygen content by about one-third, causes a 35% increase in cerebral blood flow, despite a fall in systemic blood pressure and a

decreased arterial Pco_2 and pH (due to increased pulmonary ventilation) (Kety and Schmidt, 1948) The net effect of this degree of hypoxia on cerebral vascular resistance equals that produced by the inhalation of 5 to 7% CO_2 Inhalation of high concentrations of oxygen which increase arterial Po_2, causes a reduction of cerebral blood flow, but this does not occur if care is taken to prevent concomitant hypocapnia (Turner et al , 1957)

If, as it may seem plausible, hypercapnia and hypoxia act synergistically on cerebral vascular resistance, possibly in combination with other metabolic products resulting from a diminished blood supply, this could well explain the mechanism of autoregulation exhibited by the cerebral circulation Any impairment of blood supply resulting from a drop in perfusion pressure would automatically lead to an accumulation of metabolic products which, when present in sufficient concentrations, would act to restore blood flow by producing vasodilation At the present time, however, this is still an unproved hypothesis

REGULATION OF CEREBRAL BLOOD FLOW

Whatever is the exact mechanism of cerebral vasomotor regulation, cerebral blood flow shows a remarkable constancy under a variety of physiological conditions studied so far In spite of the increased cardiac output and arterial blood pressure associated with physical exercise, cerebral blood flow remains unchanged (Hedlund et al , 1962), there may be a slight increase in cerebral vascular resistance, probably due to the concomitant drop in arterial Pco_2 secondary to hyperventilation (Kleinerman and Sancetta, 1955) Intellectual efforts, such as mental arithmetics, have no effect on cerebral blood flow (Sokoloff et al , 1955) Anxiety also seems to be devoid of influences on cerebral hemodynamics (Scheinberg and Stead, 1949) Natural sleep induces cerebral vasodilation and a small increase in blood flow without changes in arterial oxygen or carbon dioxide tensions (Mangold et al , 1955). Eating does not affect cerebral blood flow (Rowe et al , 1959) The effects of tilting from the supine to the upright position depend on the degree of tilting Moderate elevations (20° from horizontal) have no effect on cerebral blood flow (Shenkin et al , 1948), but more near-vertical positions which lower the mean arterial pressure at head level below 60 mm Hg cause, as it might be expected, a reduction in blood flow despite a decrease in cerebral vascular resistance (Scheinberg and Stead, 1949)

Cutaneous Circulation

ANATOMY

In the skin the arterioles, upon approaching the bases of the papillae (the layer of the corium immediately underlying the epidermis), turn horizontally, and give rise to metaterioles from which originate, in turn, hairpin-shaped endothelial tubes—the *capillary loops* The proximal or arterial limb of the capillary loop ascends in the papilla and then turns upon itself to form the venous limb The latter on reaching the base of the papilla joins with the venous limbs of neighboring loops to form a *collecting venule* The collecting venules anastomose with one another to form a rich plexus—the *subpapillary venous* plexus—which runs horizontally beneath the bases of the papillae and drains into deeper veins The capillary loops can be seen readily in the living skin under the low power of the microscope The vessels at the base of the human fingernail are shown diagramatically in Fig 3 173

The *color* of the skin is dependent upon the capillary loops and the subpapillary venous plexus The vessels of the plexus, though more deeply placed, present a greater area parallel to the skin, whereas the capillary loops are disposed chiefly at right angles to the skin surface When little blood is contained in the

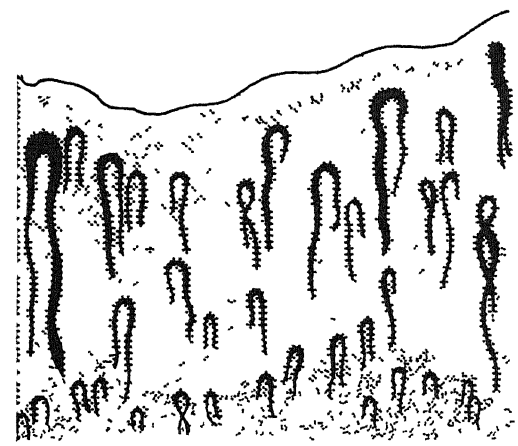


Fig 3 173 The bed of the finger nail in a healthy subject showing the capillary loops and the summits of the skin papillae (After Lewis T 1927)

superficial vessels, the skin is unusually pale and more transparent, and the deeper venous plexuses then contribute largely to the color of the skin, often adding a leaden tint to the pallor When the overlying vessels are open and the skin is well supplied with blood, these deeper vessels are hidden from view

The *hue* of the skin, i e , the dominance of the reddish or of the bluish hue, depends upon the extent to which the oxyhemoglobin becomes reduced during the passage of the blood through the cutaneous vessels The degree of reduction will depend, as a rule, upon the rate of blood flow With rapid flow, the blood is more arterial in character, with slow flow, more venous, and the color of the skin varies accordingly

CUTANEOUS BLOOD FLOW IN THE HAND AND FOOT

Only a vasoconstrictor nerve supply with adrenergic sympathetic fibers has been identified in the hand or foot, its contribution to the regulation of body temperature has been shown to be of importance The demonstration of specific cholinesterases by histochemical methods in the digital arteriovenous anastomoses suggests they may be cholinergically controlled It has also been postulated that the sympathetic nerve supply to human fingers exerts a tonic dilator influence on vessels through the release of acetylcholine The indirect evidence for this theory includes the dilation of vessels by extracts of human digital arteries, prevention of digital vessel dilation by atropine, and the presence in digital vessels of pseudocholinesterase activity

The vasodilator nonapeptide, *bradykinin*, was suggested as a mediator of cutaneous vasodilator effects associated with sweat gland activity, but the importance of this mechanism is now considered doubtful (Kellermeyer and Graham, 1968)

Body warming elicits a reflex vasodilation in the hand or foot which has been shown to correlate with a release of vasoconstrictor tone (a decreased number of impulses traveling over the sympathetic nerve fibers) The reflex may originate either in cutaneous receptors or by central nervous system stimulation The vessels themselves are also sensitive to warm temperature for the dilation following vasomotor nervous inhibition (by blocking agents) can be augmented by local heating of the hand

Arteriovenous (AV) anastomoses are communications between smaller arteries and arterioles and the corresponding venous channels, through which the blood may be shunted and capillary areas short circuited (Fig 3 174) These AV anastomoses do not appear to be under metabolic control but, rather, governed chiefly by reflex influences from temperature receptors and from central nervous system centers

CUTANEOUS VEINS

In dogs and cats, it has been shown that the superficial large and small veins of the skin have a nerve supply, undergo rhythmic changes in diameter, and respond to pharmacological agents The small veins may contract separately and independently from the large veins or arterial vessels during sympathetic nerve stimulation When the veins constrict, an increase in pressure results Thus, a reduced blood flow may occur through the vascular bed without an obstruction or constriction of the arterial tree being present In humans, forearm cutaneous vein segments, which were isolated in situ by pressure, have been demonstrated to constrict following a variety of normal and noxious stimuli The venoconstriction was blocked by chemical or surgical sympathectomy or infiltration of an anesthetic solution around the vein Venous tone was decreased during vasovagal syncope and by stroking the skin over the vein The fact that cutaneous veins constrict by cooling and are not affected by warming is evidence for the existence of only a vasoconstrictor nervous supply

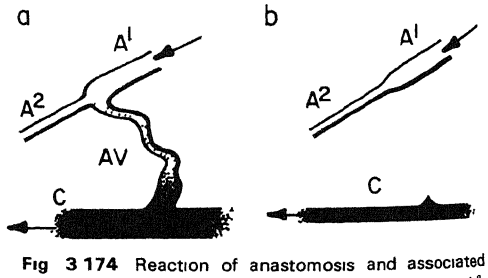


Fig 3 174 Reaction of anastomosis and associated vessels to lowering of body temperature A^1 artery, A^2 arteriole AV, arteriovenous anastomosis C vein Left figure (*a*) anastomosis open right figure (*b*) anastomosis closed (After Grant 1930)

Although active capillary constriction and dilatation has been denied, some obscure change in endothelial cells resulting in capillary dilatation has not been entirely ruled out Because skin color is thought to be governed largely by the amount of blood in venules, skin color responses are probably attributable chiefly to venular responses, but capillary dilatation is also described Therefore, in the following descriptions the term venular dilatation will be employed for changes not known to be passive effects of arteriolar or precapillary sphincter relaxation, but which may be accompanied by capillary dilatation

VASCULAR RESPONSES OF SKIN TO STIMULATION BY MECHANICAL AND OTHER AGENCIES

The White Reaction If the surface of the skin is stroked lightly with a blunt "pointed" instrument, a line of pallor appears in 15 to 20 sec which traces the path taken by the instrument The line attains its maximal intensity in ½ to 1 min , and then gradually fades to disappear in 3 to 5 min The white reaction proper is due to direct stimulation of vessel walls and has no nervous basis It has been shown to be due to the tension exerted upon the walls of the minute vessels—collecting venules and especially of the subpapillary venous plexus—which respond to the stimulus by contraction The sharply delineated character of the white line, and the fact that it can be obtained after the circulation through the region has been occluded by compression of the larger vessels, show that it is an active venular response, and not the result of arteriolar constriction

The Triple Response. This comprises *a) the red reaction, b) the flare*, and *c) the wheal*

The Red Reaction If the pointed instrument is drawn more firmly across the skin, especially of the forearm or back, a red instead of a white band appears after a somewhat shorter latent period (3 to 15 sec.), reaches its maximum in ½ to 1 min , and then gradually fades Like the white reaction it is strictly localized to the line of stroke, it is due to *dilation* of the venular vessels The red reaction can be induced in its full intensity in the skin from which the circulation has been occluded, so it is due to active dilation of the venular vessels and not merely a passive result of arteriolar dilation The red reaction is not dependent upon nervous mechanisms since it occurs after section and degeneration of the cutaneous nerves

The Spreading Flush or Flare If the stimulus is unusually strong, or is repeated often enough, the reddening of the skin is not confined to the line of stroke but surrounds it for a variable distance (1 to 10 cm) according to the intensity of the injury inflicted The temperature in the suffused area is definitely raised This flare reaction appears a few seconds (15 to 30) after the local red line, and fades sooner It is due to dilation of the arterioles and venules since it does not appear after the circulation of the part has been occluded by means of a tourniquet, also, unlike the red reaction, the flare is dependent upon local nervous mechanisms (axon reflex) It occurs after the nerves are divided but not after they have degenerated

Local Edema or Wheal When the stimulus is still more intense, the skin along the line of the injury becomes blanched and raised above the surrounding area to a height of 1 or 2 mm or even more Such a wheal or welt can be produced in a normal person by the lash of a whip and other types of strong localized stimulation In susceptible individuals, even light stimulation, such as drawing a pencil with moderate pressure over the skin of the back, will produce linear wheals surrounded by a diffuse red halo along the pencil's track In this way letters or other designs may be embossed upon the skin (Fig 3 175) This phenomenon is spoken of as *dermographism or factitious urticaria* The wheal is preceded by, but then completely replaces, the usual red reaction It makes its appearance in 1 to 3 min from the time of injury and is at its maximum height in



Fig 3 175 Dermographism (Ormsby and Montgomery 1954)

3 to 5 min It is surrounded by the flare described above The raised patch at first is clearly demarcated, but as time passes it increases in width and decreases in height, loses its sharpness and finally, although perhaps not for some hours, disappears The wheal is due to the transudation of fluid from the minute vessels involved previously in the red reaction, it is, therefore, a localized edema Increased permeability of the capillary wall is judged to be the immediate cause Wheal production does not depend upon a nervous mechanism

The triple response has been attributed to the release of some diffusible substance by the injured cells Injection of histamine produces a similar reaction, but whether histamine or some other vasoactive substance (e g , ATP, bradykinin, etc) is involved is not known

DORSAL ROOT VASODILATOR FIBERS

The dorsal root of the spinal cord contains vasodilatory fibers supplying mostly the cutaneous vessels They do not convey impulses to or from the higher central nervous system centers for control of the cutaneous circulation No true cutaneous vasomotor reflexes are seen after sympathectomy Experiments have shown that the dorsal root vasodilator fibers do not engage in the baro- or chemoreceptor control of vascular tone, nor in the regulation of heat loss from the skin The transmitter substance of these fibers is unknown but they probably act in an axon reflex arrangement to affect regional blood vessels These are probably the fibers that, upon damage to the superficial tissues, are involved in the vasodilator (flare) reaction of the triple response Any stimulus causing damage to the skin, such as severe cooling and heating, activates these fibers

CENTRAL NEURAL CONTROL OF CUTANEOUS AND MUSCLE BLOOD FLOW

The central neural control of blood flow is through the sympathetic fibers

A variety of stimuli affecting peripheral receptors elicit spinal vasomotor reflexes, the efferent limbs are the cutaneous sympathetic vasoconstrictor fibers These reflexes are usually segmentally or regionally arranged Also, in certain animal preparations, the preganglionic sympathetic neurons in the lateral horns of the spinal medulla may exhibit spontaneous activity independent of afferent stimuli

The vasomotor center located in the oblongate medulla is believed to be the primary controlling factor of cutaneous and muscle vessel vasoconstrictor tone Experiments on dogs and cats have shown that stimulation of the midline of the oblongate medulla leads to vasodilation in both skin and muscle This vasodilatory response is not blocked by atropine and, therefore, is secondary to a release of vasoconstrictor tone Stimulation lateral to this vasodilator area elicits vasoconstriction in the same regions Both these responses are absent in sympathectomized animals The vasodilator and vasoconstrictor areas may act as an integrated unit and, evidently, are primarily involved in vascular reflexes arising from the baroreceptors in the carotid sinus and aorta A decrease in blood pressure causes reflex vasoconstriction in the dog hindlimb via the carotid sinus baroreceptors while stimulation of the carotid sinus nerve induces vasodilation in skin and muscle

The vasomotor center possesses automaticity for it continues discharging impulses without apparent afferent stimuli from the periphery or higher autonomic centers The marked fluctuations from moment to moment that occur in the digital and forearm arteries and veins, and that disappear following sympathectomy, are thought to be secondary to the activity of the vasomotor center The pat

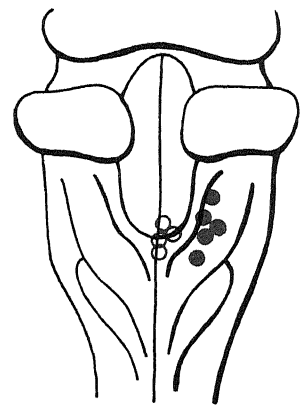


Fig 3 176 Schematic drawing of the oblongate medulla showing the situation of points yielding skeletal muscle vasodilation secondary to inhibition of vasoconstrictor tone (*open circles*) and points yielding sympathetic vasodilator discharges (*black dots*) (From Lindgren and Uvnas 1955)

terns of these continuous variations in pulse volume of the fingers, toes and ears, appear to correlate with the personality types (emotional stability or lability) of the humans studied

Stimulation of a more lateral area to the vasomotor center invokes a vasodilation in skeletal muscle through the activity of the cholinergic (blocked by atropine) sympathetic fibers (Fig 3 176) The vasodilatory area of the vasomotor center has no anatomical or apparent functional connection with these cholinergic fibers which pass through, but do not synapse in, the oblongate medulla

The anterior hypothalamus is often referred to as the "heat loss center for it controls the vasoconstrictor fibers in association with the vasomotor center Local heat applied to this area inhibits vasoconstrictor fiber discharge (vasodilation) while electrical stimulation excites the fibers (vasoconstriction) This center is thought to predominantly regulate the cutaneous vessels and, perhaps, the arteriovenous anastomoses Sympathectomy also abolishes these effects of the hypothalamus Stimulation of certain areas in the hypothalamus may elicit a vasodilation in muscle via the cholinergic fibers which synapse in this area This muscle vasodilation is often associated with a vasoconstriction of the skin

Many areas of the cerebral cortex have been found that influence, by excitation or inhibition, vasoconstrictor fiber discharge For example, stimulation of the motor or premotor cortical areas results in cutaneous vasoconstriction A muscle vasodilation also occurs but is evidently secondary to a baroreceptor reflex from a rise in blood pressure The cholinergic vasodilator fibers to skeletal muscle are also represented near the motor cortex These fibers, which probably are distributed only to muscle, synapse in the hypothalamus and mesencephalon and pass through the oblongate medulla as described above Their function in man is not completely understood, but they appear to be involved in the control of the blood flow needs of muscle and not in the regulation of blood pressure It is probable that the sympathetic vasodilator innervation of muscle is activated by emotional reactions and prepares for vigorous muscular effort

Skeletal Muscle Circulation

Resting forearm skeletal muscle blood flow has been estimated to range normally from 1 8 to 9 6 ml per 100 ml of muscle per min The blood flow in skeletal muscle is affected and controlled by a multitude of factors Many of these factors are poorly understood and their relative importance in the normal functioning of skeletal muscle is not known Investigation of skeletal muscle blood flow is performed in man by techniques similar to those outlined under skin blood flow The plethysmograph measures both the skin and muscle blood flow, the forearm is considered representative of muscle and the hand or foot, of skin because of the relative amounts of each tissue in these regions However, it has been shown that 60% of the total forearm flow may be in the skin under certain conditions The skin may be iontophoresed with epinephrine to limit or abolish its blood flow in order to obtain forearm muscle flow determinations However the depth of penetration of the epinephrine is a problem and may alter the muscle blood flow In animals blood flow may be measured by more direct methods and the results of central nervous system manipulations may be studied

NEURAL CONTROL OF SKELETAL MUSCLE BLOOD FLOW

Vasoconstrictor fibers have been shown to supply skeletal muscle by the demonstration of an increased blood flow in human forearms following nerve blocks, and by other indirect procedures The transmitter substance for these constrictor fibers is presumably norepinephrine Intra-arterial epinephrine has both a vasodilator and vasoconstrictor effect (see α, β, and γ receptors, p 3-171) on human skeletal muscle vessels, while norepinephrine has only a vasoconstrictor action Vasodilator cholinergic sympathetic fibers also supply skeletal muscle as demonstrated by a responsiveness of the bed to acetylcholine, blocking of the vasodilator response with atropine, and the results of studies performed during vasovagal syncope Despite a fall in blood pressure in vasovagal syncope, the forearm blood flow increases (and hand flow decreases) If the patient has had a sympathectomy, the forearm blood flow does not increase demonstrating that an active vasodilation via sympathetic nerve fibers had presumably occurred in the intact extremity These vasodilator fibers appear to be activated by higher brain centers as described previously and do not participate in reflex blood pressure adjustments

At present, the vasoconstrictor fibers are believed to be primarily concerned in postural and other homeostatic blood pressure adjustments When the legs of a subject are raised, an increased blood flow occurs in the forearm but not the hand This reflex has been shown to be secondary to an increased volume of blood in the thorax and an elevated central venous pressure which probably stimulates receptors in the low pressure vascular bed of the thorax Since atropine does not abolish the forearm vasodilation, it is considered to be a result of a release of vasoconstrictor tone

BLOOD GASES, pH AND IONS IN SKELETAL MUSCLE BLOOD FLOW

The effect of blood gases and pH on skeletal muscle blood flow has been referred to in Chapter 8 Studies in the dog forelimb show that many factors affect the vascular bed Arteriolar dilation is elicited by increased blood levels of potassium, magnesium, acetate, citrate, pyruvate, AMP, ADP, ATP, hydrogen ions, increased plasma tonicity and decreased calcium blood levels A decrease in hydrogen ions causes arteriolar constriction, elevated calcium concentrations or decreased plasma tonicity act similarly Combinations of low potassium, increased calcium, low magnesium, and alkalosis produce more vasoconstriction than the ions alone Changes in blood oxygen content induce only small changes in forelimb resistance Studies on isolated vessels and in denervated, adrenergically blocked limbs demonstrate that the ions act directly on vascular smooth muscle The importance of these studies is apparent when the many disease states with electrolyte abnormalities are considered

Studies in man have shown that the increase in forearm blood flow during hyperventilation (increased pH, decreased arterial P_{CO_2}) is *1)* prevented by breathing air mixtures containing 5% CO_2, 2) occurs only in muscle and not in skin, and *3)* is not abolished by blocking the nerve supply to the limb This evidence points to a humoral mechanism which causes the increased forearm blood flow by a direct action on peripheral vessels It is interesting that a similar increase in forearm blood flow is found when 5 to 10% oxygen is breathed If CO_2 is added to the latter mixture, the blood flow response is greatly attenuated Therefore the decreased peripheral resistance present in hypoxia may be secondary to hypocapnia

Inhalation of 5 to 10% CO_2 (or intravenous sodium bicarbonate) does not increase skeletal muscle blood flow unless sympathetic blockade is present Thirty percent CO_2 inhalation decreases forearm blood flow, nerve block or general anesthesia abolishes or even reverses the vasoconstriction Intra-arterial sodium bicarbonate also increases forearm blood flow Thus hypercapnia produces an *indirect* muscle vasoconstriction and a *direct* vasodilation The vasoconstriction is believed to be elicited by the effect of increased CO_2 (or decreased pH) on the carotid or aortic body chemoreceptors

A decreased pH of the blood by 0 08 units (intravenous lactic acid or ammonium chloride) or an increased pH (intravenous sodium bicarbonate) induces no change in forearm blood flow If the sympathetic nerves are blocked, an increased or decreased pH leads to an increase in blood flow, therefore a sympathetic vasoconstriction was concealing a direct vasodilation due to the local pH (or CO_2) effect

In the above studies, the separation of the effects of arterial P_{CO_2} and pH on the peripheral blood vessels is difficult In humans, local increase in P_{CO_2} or pH changes in either direction from normal produce vasodilation Systemic increases in P_{CO_2} or changes in pH in either direction induce sympathetic vasoconstriction, systemic hypocapnia causes vasodilation Data obtained in the dog hind limb does not entirely agree with these findings, the local effect of an increased hydrogen ion concentration is arteriolar dilation but a lowered pH caused by an increased P_{CO_2} has little effect Before it can be decided whether pH or P_{CO_2} is the important variable, further study is necessary It must be remembered that changing the pH of blood also affects P_{CO_2}

MECHANICAL EFFECTS ON SKELETAL MUSCLE BLOOD FLOW

When perfusion pressure is plotted against skeletal muscle blood flow (as determined in isolated muscle beds of the hindlimb of a dog), a parabolic-shaped curve is found in the upper and lower ranges (Fig 3 177), as in the skin vascular bed These two normal relationships of pressure to flow are joined in the intermediate range by a sigmoid curve (a large drop in

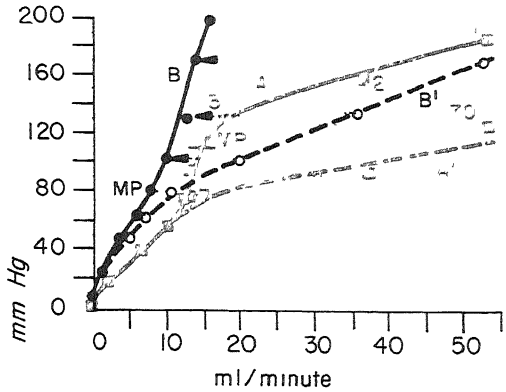


Fig 3 177 Plots of relationship between perfusion pressure and blood flow in the portion of the canine quadriceps muscle supplied by the branches of the femoral artery between the inferior epigastric and saphenous branches Plots *A* and *B* were obtained 6 hr apart in the same animal A^1 and B^1 (*dashed lines*) flow during first 2 sec of perfusion at each pressure *A* and *B* (*solid lines*) stabilized flows after maintenance of perfusion pressure for 10 to 20 sec Arrows indicate change of flow with time at each perfusion pressure MP represents the level of mean arterial pressure Ordinate scale represents the perfusion pressure in mm Hg expressed as the difference between arterial and venous pressures abscissal scale blood flow in milliliters per minute (from Green et al 1944)

perfusion pressure changes blood flow little) It is postulated that the reason for the sigmoid curve is a reactive dilation at perfusion pressures below mean normal aortic pressure to maintain blood flow to the muscular bed The two parabolic areas on the curve would represent perfusion pressures during two states of vasomotor activity of the vessels Because of the shape of the curve, changes in vasomotor activity, secondary to a variable being tested, cannot always be predicted from the perfusion pressure and blood flow

METABOLITES AND SKELETAL MUSCLE BLOOD FLOW

The skeletal muscle vascular bed is subjected to very active and widely ranging fluctuations in blood flow The idea has often been expressed that the formation of metabolites by the muscle cells is the main factor in controlling the bed Since direct stimulation of an animal's curarized muscle still elicits an increased blood flow, while motor nerve excitation does not, it is felt that the hyperemia of contracting muscle is secondary to a local physicochemical change. Also, humans with sympathectomized extremities can perform as well as those with an intact innervation Local or reflex warming or cooling has little effect on this intrinsic vasomotor tone of muscles It has been suggested that the metabolites, accumulated from the muscle cell metabolism, may act directly on the smooth muscle cells of vessels or indirectly via a local nervous mechanism, such as peripheral ganglion cells or an axon reflex The existence of ganglion cells among the muscle fibers has not been proven

In the dog muscular activity induces a dilation of the femoral artery which is mediated by an axon reflex (elicitable by acetylcholine) This mechanism, if it was shown to be present in the smaller arterial branches, could account for the vasodilation of exercise In the cat limb, post exercise muscle hyperemia is blocked by substances (cocaine, botulinum toxin) which are likely to abolish such axon reflexes The vascular reactivity is not paralyzed by these agents for the vessels continue to respond to epinephrine and acetylcholine Since botulinum toxin specifically paralyzes cholinergic nerve endings, the fibers in the axon reflex may be cholinergic However, a strong argument against this theory is that the vasodilation of exercise hyperemia is not blocked by atropine (although atropine does not block cholinergic fibers in all areas)

The role that local metabolites play in stimulating such axon reflexes is not known Anoxia, increased CO_2 tension, lactic acid, hydrogen ions, bradykinin, histamine, acetylcholine, adenosine triphosphate, adenylic acid, and potassium ions have been suggested as the local determinant eliciting exercise vasodilation Most of the metabolites have been studied by infusion into the musuclar vascular bed or by examination of venous blood leaving the muscle following exercise hyperemia Each one has received little support as the dilator agent, but a combination of two or more of these factors may provide the answer

Bradykinin has been shown to be a very potent vasodilator in skeletal muscle by plethysmography and radioisotope disappearance rates No change in oxygen consumption occurs despite the large increase in human calf blood flow However there is no evidence for the presence of bradykinin activity in skeletal muscle, the venous blood draining exercising muscle does not contain increased bradykinin

activity These results do not rule out bradykinin as an agent affecting skeletal muscle blood flow, since the half-life of bradykinin in the blood is extremely short and only biological assay methods are available for its detection

The circulation in human skeletal muscle has been studied by indirect methods during sustained muscular contraction It was found that below a certain critical strength of sustained contraction of the calf muscle, blood flow was not decreased, but, above this critical point, the vessels were evidently compressed, for hyperemia followed release of the contraction

With rhythmic contraction of calf or forearm muscles, blood flow was increased during and immediately following exercise It was felt that increased flow (vasodilator metabolite mediated) alternating with decreased flow (mechanical compression of vessels) paralleled the relaxation and contraction of the muscle Mechanical obstruction, metabolite formation, and the action of muscles on the venous flow are probably all intricately involved in the muscular blood flow of exercise It has been shown that an increase in venous pressure may alter arterial inflow to muscle The sympathetic nervous system probably does not have a significant effect on the circulatory changes during exercise, for vasodilation is similar during the after exercise in normal and sympathectomized limbs Also, the vascular bed of the exercising hindlimb of the dog becomes less responsive to sympathetic stimulation as the strength of contraction and oxygen uptake increases

The blood flow of exercising skeletal muscle in the dog, except during maximal exercise, is reduced by sympathetic nerve stimulation intra-arterial norepinephrine or epinephrine infusions, and carotid sinus stimulation However the hypothalamic vasodilator sympathetic pathway does not affect exercising muscle Both resting and active skeletal muscle show autoregulation in the dog An increased resistance occurs with a rise in perfusion pressure and vice versa, i e despite changes in blood pressure, the blood flow tends to remain stable Theories to explain autoregulation involve a myogenic responses of the vessel walls, release of vasodilator metabolites, or decreased oxygen The level of metabolic activity does not affect autoregulation, the phenomenon also occurs in denervated muscles

CIRCULATORY ADJUSTMENTS WITH EXERCISE

The changes in regional blood flow with exercise are summarized in Fig 3 178 As expected, the increased cardiac output supplies an enormously elevated skeletal muscular blood flow Skin flow increases at first, until 50 to 60% of the maximal work load is reached, then declines This increase favors body cooling But further increase in work load causes a progressive decrease in skin flow as sympathetic vasoconstrictor tone continues to rise, apparently overcoming thermoregulatory skin vasodilator responses Visceral flow (splanchnic, hepatic, and renal) decreases, cerebral flow remains unchanged, and coronary flow increases in proportion to cardiac work These findings are, in general, based on indirect determinations of blood flow applied one at a time to various vascular beds, e g , renal clearance (sect 5, chap 1) hepatic bromsulphalein (BSP) excretion, skin temperature changes, etc (see also p 3-142) However, in dogs with implanted flowmeters, renal artery flow remain unchanged while hepatic artery flow to liver and stomach and through hepatic vein decreased with short term treadmill exercise (Rushmer, 1970) Such responses appear

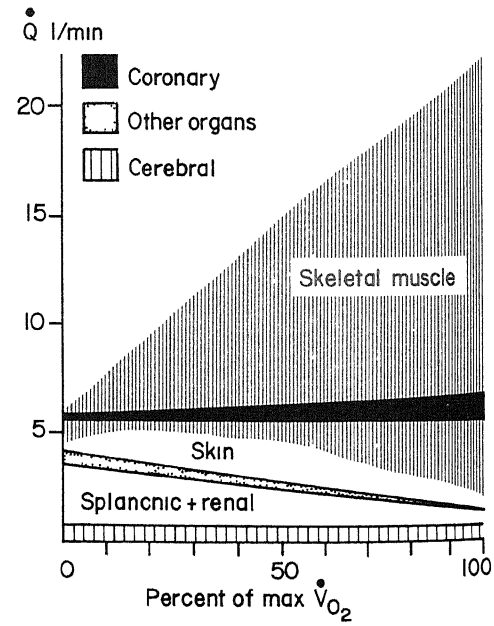


Fig 3 178 The changes of regional distribution of cardiac output as work load is increased Q̇ blood flow V̇O2 oxygen consumption as a measure of work load (Redrawn from Clausen 1969)

characteristic of short term work and brief bouts of exhausting exercise In prolonged moderate work or sustained strenous exercise, blood flow distribution to kidney, liver, gastrointestinal tract, and skin must be maintained at levels adequate for continued functions of these organs Telemetered flowmeter data (mesenteric artery, renal artery) from sledge dogs indicate no decrease in splanchnic and renal flow with prolonged strenuous exercise (Van Citters and Franklin 1969b) More work is required to reconcile various results based on findings from studies employing different methods and species At present, the consensus remains that with short term exercise abdominal visceral and renal flow decrease, cerebral flow is unchanged, skin flow increases with moderate, decreases with severe work, coronary flow increases with cardiac output, and flow to skeletal muscle increases During prolonged exercise, the elevated flow to skeletal muscle is not produced at the expense of adequate flow to abdominal viscera

Physical conditioning or training affects the cardiopulmonary and skeletal muscular systems in a variety of ways which improve work performance (Clausen, 1969) Stroke volume, blood volume, total hemoglobin, muscle capillary bed, and mitochondrial content of muscle are all increased Thus, all oxygen delivery and utilization systems involved in exercise are enlarged The capacity to increase cardiac output is elevated by the increase in stroke volume, which is reflected in lower heart rates at rest and for given work levels Muscle blood flow at a given work load is actually reduced (Clausen, 1969) This occurs because there is an increased oxygen extraction in trained muscle, made possible by an increase in oxidative mitochondrial enzymes and in the number and size of mitochondria Thus, "white" muscle fibers come to resemble "red" muscle fibers in histological appearance and enzymatically, an adaptation favoring aerobic metabolism and endurance (see p 3–35 and Tables 3 7 and 3 9) This change appears to be mediated by motor neuronal activity, i e , muscle composition is perhaps governed by the average rate of stimulation or migration of a specific trophic substance from the central nervous system, a hypothesis supported by cross innervation experiments which can cause transformation of white fibers to red and vice versa (Guth, 1968, Buller et al , 1969)

The net effect of these changes is that following physical training at a given work load myocardial and skeletal muscle blood flow can actually be reduced and less reduction in splanchnic blood flow is required for adequate muscle perfusion (Clausen, 1969) This may explain the failure to record reduction in mesenteric and renal blood flow in racing sledge dogs, as contrasted with reported decreases in splanchnic flow in exercising men not trained for endurance performances On the other hand, the microcirculatory bed to muscle is enlarged, so that greater peak flows can be attained

Thus, physical conditioning is a complex affair in which cardiac performance is improved through increase in stroke volume, reduction in heart rate, and decrease in peripheral resistance, peak muscle blood flow can be increased by enlargement of the microcirculatory bed, but the demand for muscle blood flow is relatively decreased by elevation of oxygen extraction from blood

REACTIVE HYPEREMIA

Following a period of complete occlusion of the arterial supply to a limb, the blood flow increases markedly and then returns to the control level This phenomenon has been named "reactive hyperemia " For measurement of reactive hyperemia, the blood flow in excess of that expected (the control blood flow) is usually calculated The response occurs both in skin and muscle and can be decreased in the skin by cooling, epinephrine, or tobacco smoking The amount and duration of reactive hyperemia correlate with the previous length of arterial occlusion This has been shown to be true for occlusion periods up to 10 min

The blood flow and oxygen debt incurred during the period of occlusion are calculated by multiplying the control blood flow or oxygen usage before circulatory arrest by the duration of occlusion This assumes that the blood flow and oxygen usage of the muscle would have remained at the previous control rates if the blood flow had not been occluded and that the metabolic rate of the muscle is not affected by the circulatory arrest In the isolated gracilis muscle of the dog and in the human forearm, the oxygen debt is approximately repaid for different periods of occlusion, while the blood flow debt repayment ranges from 50 to 200% of that expected

The phenomenon of reactive hyperemia has also been demonstrated in the exercising dog

hindlimb Flow debts are underpaid or barely repaid Oxygen debts are entirely or partially repaid depending upon the level of muscle performance and the duration of the ischemic period In resting muscle and during light exercise, both increased blood flow and oxygen extraction are involved in repayment of the oxygen debts but during medium and strong exercise, the increased blood flow is most important Since oxygen debts are often not repaid during strong exercise in the presence of a *decreased* arteriovenous oxygen difference, oxygen extraction is the limiting factor

The metabolites incriminated as the cause of exercise hyperemia have also been postulated as eliciting reactive hyperemia Since reactive hyperemia occurs in sympathectomized and denervated limbs, nervous system control is not considered as important The mechanical effect of the lack of pressure in the vessels during the period of occlusion has been demonstrated to play a role in the dilation If the forearm is "packed" with venous blood before the circulatory arrest, the ensuing reactive hyperemia is decreased However, the peak flow of the reactive hyperemia is not altered so that intravascular pressure cannot be the entire cause of the dilation Lewis and Grant (1925) postulated that H-substance was involved in the production of reactive hyperemia, some observers have found an increased histamine concentration in venous blood after circulatory arrest but others have not However, antihistamines do not decrease nor abolish reactive hyperemia As in exercise hyperemia the vasodilator metabolites and effects of pH cannot be ruled out by studying only the venous blood, tissue studies would probably be more informative Failure to discover the causative agent of vasodilation in muscular exercise and following arterial occlusion has been a major obstacle to the understanding of muscle physiology and the diseases involving a decreased blood supply to human limbs It seems likely that the study of aerobic and anaerobic tissue metabolism of skeletal muscle would be more profitable than blood flow studies, which evidently do not correlate with oxygen usage

ARTERIOVENOUS ANASTOMOSES IN SKELETAL MUSCLE

The question of whether arteriovenous shunts actually exist as functional units in skeletal muscle is still unsettled to most physiologists Microscopic studies on the circulation of rat skeletal muscle reveal many communications between small arteries and veins which do not enter the muscle proper During inactivity, most of the flow is through these anastomoses and not to the muscle fibers When the rats are bled in small amounts, the muscular arterioles close and flow in the capillaries ceases while it persists in the anastomotic channels However, that such communications actually function in man and possess a means of shunt regulation has not been demonstrated Indirect evidence for the existence of functional anastomoses is the decreased oxygen consumption and increased lactic acid production during vasodilation, such findings would occur if capillary flow were reduced As mentioned previously, the disappearance rate of a radioisotope from a depot in skeletal muscle is thought to represent only capillary blood flow and not shunt flow For example, during emotional stress, total forearm blood flow increases (plethysmograph) but radioisotope clearance from forearm muscle does not change Such studies have led to the concept of a dual circulation in *resting* muscle, one circulatory pathway has been referred to as "nutritional" and the other, "nonnutritional "

COLLATERAL BLOOD FLOW

When the major artery to a vascular bed is occluded, smaller vessels arising from arteries above the occlusion supply the area with blood This collateral flow may prove adequate for all functions Some patients withstand an embolic femoral artery occlusion with no adverse effects to the limb In dogs, ligation of the femoral artery rarely produces signs of ischemia Since collateral arteries function within seconds of arterial occlusion, it is thought that these vessels pre-exist, however they increase in size following arterial occlusion In patients with aortic bifurcation thromboses, aortography often reveals vessels as large as the femoral artery arising from the mesenteric arteries and supplying the limbs

In normal young subjects, plethysmographic studies of calf blood flow show that mechanical compression of the femoral artery decreases flow to one-sixth of the previous resting level with recovery to normal flow within 6 min despite continued compression The flow is non-, or only slightly, pulsatile Reflex warming of the body and ganglionic blockade

increase the collateral blood flow demonstrating that it is affected by the sympathetic nervous system Exercise of the leg usually increases the collateral flow but the increase may be due to a rise in systemic blood pressure

When the external iliac or femoral artery is acutely occluded in dogs or cats, the flow drops to one-third of the previous resting level and then gradually rises to 50 to 100% of normal In dogs with acute or chronic arterial occlusions, the blood pressure below the obstruction initially falls to low levels, but approaches normal in a few weeks

11 / SPLANCHNIC, RENAL, AND FETAL CIRCULATIONS

Splanchnic Circulation

The splanchnic circulation may be represented as consisting of three parts *1*) the mesenteric bed, supplying the gastrointestinal tract, *2*) the splenic bed, and *3*) the hepatic bed

A unique feature of this circulatory system is that the combined outflow from two of the components (mesenteric and splenic) constitutes the major portion of the inflow of the third one (hepatic), through the portal vein In its hemodynamic implications, this anatomical arrangement has been compared to the Wheatstone bridge in an electrical circuit (Bradley, 1958), because of the placement of the resistance vessels (Fig 3 179) This serves mainly to emphasize the large number of variables that influence blood pressure and flow in any one point of the system, and the potential fallacy of conclusions based upon measurements of a single variable, such as portal venous pressure

As in other regions of the circulation, it is the measurement of blood flow that presents the greatest source of difficulties Direct methods to this date can only be applied at the cost of considerable surgical and anesthetic trauma, and with various degrees of interference to flow (particularly venous flow) due to cannulation procedures This latter obstacle has been eliminated by the application of noncannulating electromagnetic flow meters (Green et al , 1959) Indirect methods, based on the Fick principle and the ability of the liver to clear certain dyes, such as bromsulphalein (Bradley et al , 1945) or colloidal substances from the blood stream, do not have these disadvantages, but—aside from their failure to meet strictly the criteria of the Fick principle—can be used only for measurement of total splanchnic outflow (i e , hepatic venous flow), not for that of its various components

MESENTERIC CIRCULATION

The nature of pressure-flow relationships in the intestinal vascular bed is controversial at the present time According to one group of investigators, the mesenteric bed behaves in a passive manner, similar to other areas such as the skin There exists a "critical closure" pressure of the order of 15 mm Hg below which flow ceases, above this point, progressive increments of pressure result in greater and greater increments of flow in other words, vascular resistance decreases with increasing perfusion pressure (Selkurt and Johnson, 1958, Hinshaw, 1962) The reasons for this deviation from Poiseuille's law are dilatation of blood vessels with increased intraluminal pressure and the opening up of "*dormant*" capillaries This view is opposed by others who find evidence for mesenteric vascular auto-regulation, such as is seen in the brain and kidney—i e , a tendency to maintain blood flow (or possibly another variant, such as capillary pressure) constant in the face of a varying perfusion pressure This implies an increase in vascular resistance as perfusion pressure is increased (Johnson, 1960, Texter et al , 1962, Folkow et al , 1963) In all likelihood, differences in experimental methods, particularly in perfusion techniques, are responsible for the conflicting results

Certain hemodynamic characteristics peculiar to the mesenteric circuit derive from the location of these vessels in the walls of viscera capable of actively or passively changing their intramural tension Isolated, perfused intestinal loops exhibit two sorts of spontaneous contractile activity rhythmic segmental contractions, and alterations of the muscular tone Each of these has a definite influence on blood flow within the loop (Sidky and Bean, 1958) During segmental contractions, there is an increase in venous outflow and decrease in arteriolar inflow, porportional to the strength of the contraction, reverse changes occur during relaxation In addition to these phasic changes associated with individual contractions, there is close relationship between mean blood flow per unit time and the frequency of rhythmic segmental activity All other things being equal, the minute volume flow increases with

increasing frequency of contractions Changes in the gut's muscular tone are of longer duration and involve greater lengths of intestine than segmental contractions An increase of tonus results in reduction of both arterial inflow and venous outflow Following cessation of tonic contraction, blood flow temporarily increases above precontraction levels, in a manner analogous to the reactive hyperemia observed in skin, skeletal muscle and myocardium (see Chap 10) Since segmental and tonic contractions occur independently from one another and may be more or less superimposed, their effect on intestinal blood flow can be quite complex If these processes observed in the isolated loop also operate in the intact animal or man, the musculature of the intestine may emerge as a factor in the regulation of splanchnic and systemic blood flow

Passive changes in intramural pressure, such as those occurring in acute distension, also influence intestinal circulation (Noer et al , 1951) It was shown in serosal vessels of rabbit's intestine, that at intraluminal pressures of 15 to 20 mm Hg, segmental contractions cease, accompanied by slowing of blood flow in the venules With progressive rising of the pressure, circulation slows, then ceases, first on the venous, then on the arterial side While such levels of intraintestinal pressures are clearly outside the physiological range, these circulatory changes are the basis of the massive trapping of blood and leakage of plasma into the wall and lumen of the gut observed during acute intestinal obstructions

In the maintenance of normal mesenteric circulation, the hemodynamic characteristics of the local vascular bed and the possible influence thereon of the intestinal smooth musculature are supplemented by a rich network of vasomotor nerves These nerves belong to the sympathetic nervous system, their preganglionic fibers are contained in the splanchnic nerves, the postganglionic fibers form large plexuses around the abdominal aorta and accompany its branches to all the viscera The function of vasomotor nerves has been discussed in detail in Chapter 8 Electrical stimulation of the splanchnic nerves in anesthetized dogs produces sharp decline in blood flow in the mesenteric artery, along with increasing the arterial (and portal vein) pressures, indicating that vasoconstriction has occurred There is no evidence for the existence of vasodilator fibers in the splanchnic nerves (Deal and Green 1956) Intra-arterially injected norepinephrine reproduces the effect of splanchnic nerve stimulation (see chap 8), while section of the nerves or administration of ganglionic blocking agents increases mesenteric flow

Vasomotor reactions in the mesenteric circuit are of considerable importance in various circulatory adjustments (see pp 3-173 and 3-174) Not all vasomotor activity, however, is necessarily mediated by the nervous system In the dog, experimental elevation of the portal venous pressure results in a reduction of mesenteric blood flow greater than the decrease in mesenteric artery-portal vein pressure gradient, indicating that mesenteric vascular resistance had increased (Selkurt and Johnson, 1958) This change in resistance is shown to be independent from nervous activity and due to "myogenic" response of the arterioles to the pressure increase transmitted through the capillaries (Johnson, 1959) It is not known what role, if any, this venous-arteriolar response plays in the intact animal

SPLENIC CIRCULATION

The nature of the intrasplenic circulation is a long-standing, unsettled controversy Lack of agreement is centered around the manner of connection between arterial and venous capillaries whether their lumina are continuous with one another ("closed circulation") or is arterial blood discharged directly into the meshwork of the splenic pulp from which it seeps back into the venous capillaries ("open circulation") The basis for suspecting a discontinuity of the circulatory bed (unparalleled elsewhere in the body) is the presence of blood cells and intra-arterially injected foreign material in the splenic pulp, seen on histological examination, this is attributed to experimental artifacts by adherents of the "closed circulation" theory Microscopic transillumination studies in the living animal (Whipple et al , 1954) suggest that both open and closed circulatory patterns function simulatenously ("combined circulation"), but it is debatable whether these observations on mouse spleen are applicable to the human organ in view of the demonstrated structural differences between the two (Snook, 1950) Pressure-flow relationships in the perfused dog spleen show a continuous drop in vascular resistance, without any plateauing, even at high pressure and flow

rates This great distensibility of the splenic intravascular space is probably due to the opening of the vascular sinuses

In certain mammals (dog, cat, horse, guinea pig) the spleen serves as a reservoir of blood Under ordinary circumstances, blood is stored in the distensible venous sinuses, to be released into the circulation when need arises (as during exercise, anoxia, hemorrhage)

Evidence indicates that this splenic emptying mechanism is under sympathetic regulation (Green et al , 1960) Stimulation of splenic nerve fibers in the dog results in simultaneous decrease in arterial inflow, increase in venous outflow, and decrease of splenic weight The inflow response can be reproduced by intra-arterial norepinephrine injection, while the outflow response can best be reproduced by injection of epinephrine This suggests the existence of separate mechanisms for the control of splenic inflow and outflow a norepinephrine-mediated arteriolar inflow constriction and an epinephrine-mediated emptying mechanism This latter might be due either to venular relaxation causing passive elastic recoil of splenic tissues, or to primary contraction of splenic smooth muscle fibers

In man, there is no evidence to indicate that splenic contractions take place under physiological circumstances, nor do injections of adrenaline raise the circulatory plasma or red cell volumes (Parson et al , 1948) as one would expect if the spleen served as an important blood depot

HEPATIC CIRCULATION

In addition to receiving its arterial blood supply through the hepatic artery, the liver is also perfused by the mesenteric and splenic venous outflow through the portal vein As pointed out earlier, this unique circulatory arrangement leads to considerable complexity of splanchnic hemodynamics, and information regarding the various factors determining blood flow through the liver is still fragmentary The following may illustrate the difficulties encountered

The hemodynamic properties of the portal venous bed of the isolated liver show a great deal of similarity to those of other vascular beds, such as the mesenteric described earlier There is, in particular, a progressive decrease of vascular resistance with increase of perfusion pressures throughout the physiological pressure range (up to 12 to 13 cm. of H_2O) This would indicate that relatively small changes of portal pressure are associated with marked flow effects and that, consequently, portal venous pressure is an important determinant of total hepatic flow

As may be seen from Figure 3 179, however, portal venous pressure is under the influence of a variety of factors Mesenteric and splenic arteriolar tone regulate inflow, while the outflow is determined by venular tone and he patic sinusoidal pressure The latter, in turn, is dependent upon hepatic arteriolar tone and

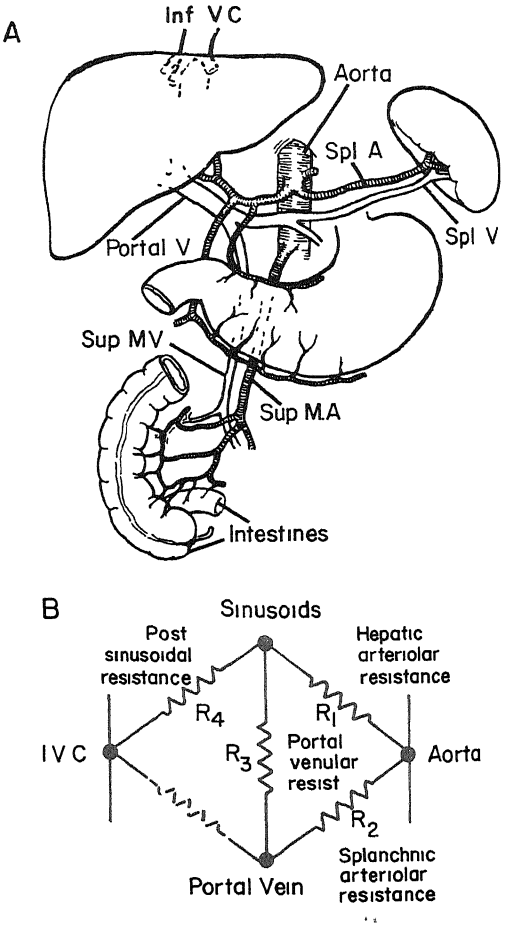


Fig 3 179 The Splanchnic circuits The vascular resistances in the splanchnic bed (*A*) are shown here in diagrammatic form (*B*) The resistances indicated are determinants of pressures in the portal vein and in the liver sinusoids and of flows through the portal vein and hepatic artery R_1 hepatic arteriolar resistance R_2 splanchnic (i e mesenteric pancreatic gastric and splenic) arteriolar resistance R_3, portal venular resistance R_4, postsinusoidal or hepatic venular resistance The dotted line represents a fifth resistance lying in the direct communications between the portal vein and inferior vena cava (From Bradley, 1958)

hepatic venous pressure (determined by inferior vena caval pressure) which regulate inflow and outflow in the sinusoids Active contractions of the portal and hepatic veins are possible additional factors all capable of altering portal venous pressure and thus influencing total hepatic flow

Experimental separation of all these variables has not been possible up to now nonetheless, certain important observations have been made

Continuous simultaneous direct measurements of hepatic arterial and portal venous flows in anesthetized dogs show the contribution of the portal vein to total hepatic flow as being approximately 65% in one study (Selkurt and Brecher, 1956), and 80% in another (Green et al, 1959) The effective portal perfusion pressures under these circumstances are 6 and 8 mm Hg, respectively Unlike the portal venous bed, the hepatic arterial system seems to possess autoregulation—i e hepatic arterial flow tends to remain constant when the perfusion pressure is varied within the physiological range (Torrance, 1961) Splanchnic nerve stimulation causes an increase in both hepatic arterial and portal venous pressures, with a concomitant reduction in the corresponding flows The calculated increases in resistance of the portal venous and hepatic arterial beds approximately equal one another, suggesting an equal degree of splanchnic vasomotor innervation Compared to the mesenteric bed, they are both less sensitive to vasomotor stimulation The effects of epinephrine and norepinephrine are similar to those of splanchnic nerve stimulation, causing, however, a less marked increase in the resistance of the portal bed Thus, even though portal vein and hepatic artery pressures increase and hepatic arterial blood flow diminishes, portal flow does not change much and may even increase (Green et al, 1959) In fact, the net effect of epinephrine in the resting, nonanesthetized human is an increase of total hepatic flow, as estimated by indirect Bromsulphalein retention (BSP) measurements, along with an elevation of portal venous pressure (Bradley et al, 1952) Norepinephrine under similar conditions causes a decrease of estimated liver flow in man (Bearn et al, 1951), while in normal, anesthetized dogs either no flow effect (Smythe et al, 1953) or a decrease in flow (Turk and Shoemaker, 1962) was observed

The hemodynamic properties of the hepatic vascular bed and the nervous and humoral controls exerted on it play an important part in circulatory homeostasis by adjusting splanchnic blood flow in response to varying body needs Blood flow through the splanchnic bed is sufficiently large to enable it to serve as a reservoir The circulating splanchnic blood volume in man at rest and in the anesthetized dog measured by isotope dilution techniques, forms approximately 20% of the total blood volume (about 1000 ml in man) (Bradley et al 1953) The total amount of blood perfusing the liver through the hepatic artery and portal vein, estimated by the BSP method amounts to approximately 1500 ml/min in man or about 20% of the resting cardiac output

CONTROL OF SPLANCHNIC BLOOD FLOW

Changes in splanchnic hemodynamics occur under various physiological and pathological conditions In man, assumption of the erect position is associated with widespread reflex vasomotor adjustments opposing the hydrostatic forces which would lower the blood pressure in the upper half of the body (see Chap 9) The splanchnic bed is an important effector end organ in these reflexes Passive tilting of human subjects from supine to upright position causes a decrease of splanchnic blood flow without similar changes of mean blood pressure, indicating that an increase in splanchnic vascular resistance has occurred (Culbertson et al, 1951) This channels blood to other vascular areas, notably those in the upper regions of the body In subjects with previous splanchnicectomy, the same tilting procedure causes arterial hypotension and a porportional decrease of splanchnic blood flow without changing the vascular resistance (Wilkins et al, 1951) This would indicate that the vasoconstriction is normally mediated through the splanchnic sympathetic nerves

Vasomotor adjustments and redistribution of circulating blood are also prominent during exercise In normal human subjects performing light exercise in the recumbent position, circulating splanchnic blood volume and splanchnic blood flow may decrease significantly (Wade et al, 1956) As exercise is usually associated with an increase in arterial blood pressure and cardiac output (see chaps 5 and 8), the reduction in splanchnic flow must be caused by increased vascular resistance The advantages for the organism of such vascular adjustments are evident, vasoconstriction diverts blood flow to

those organs that need it most (e g , the exercising muscles) Moreover, the extra blood derived from the reduction in circulating splanchnic blood volume serves to boost venous return, thereby contributing to the elevation of cardiac output (see also chap 5)

Finally, the splanchnic circulation plays an important role in the response of the organism to a reduction of circulating blood volume, such as occurs after hemorrhage This response is conditioned to a great extent by the amount of blood lost, and it may vary, accordingly, from transient vasomotor changes to a state of irreversible circulatory collapse resulting in death

It might be expected from the splanchnic responses to tilting and to exercise that blood loss would produce vasoconstriction in the splanchnic bed, in order to maintain flow in other parts of the body such as the brain and heart Actually, observations on the behavior of the mesenteric and splanchnic vasculature in response to hemorrhage are conflicting Splanchnic blood volume and flow are reduced during hemorrhage, but investigators using the BSP method for measuring splanchnic blood flow report no significant change in splanchnic resistance following removal of about one-third of the total blood volume of the anesthetized dog (Reynell et al , 1955) This would indicate that the splanchnic bed, unlike the vasculature of the skin, skeletal muscles and kidneys, does not participate in the systemic vasoconstriction following hemorrhage On the other hand, when superior mesenteric arterial flow is measured directly by means of an electromagnetic flowmeter, moderate, non-shocking hemorrhage produces an immediate increase in mesenteric vascular resistance, proportional to the magnitude of blood loss (Abel and Murphy, 1962) Removal of larger amounts of blood, resulting in irreversible shock, produces more marked but qualitatively similar changes in hemodynamic parameters The metabolic consequences, however, are quite different Splanchnic oxygen consumption is not affected by the moderate blood loss, the diminished flow is compensated for by increased oxygen extraction by the tissues (Hamrick and Myers, 1955) With more massive hemorrhage, however, blood flow is so reduced that this compensation is inadequate and splanchnic oxygen consumption falls, resulting in hepatic hypoxia. Hypoxia of the intestines and liver may be an important causative factor in determining the irreversibility of fatal hemorrhagic shock (Selkurt, 1959)

Renal Circulation

Renal circulation and renal function are so closely interrelated that reference to only one without consideration of the other always has to be artificial and incomplete Also, much of our knowledge about renal blood flow stems indirectly from the behavior of some special functions of the kidney The reader is referred to the chapter on "Urine Formation" where renal function, the anatomy and the clearance method for renal blood flow measurements are considered in detail (sect 5, chap 1)

METHODOLOGY OF RENAL BLOOD FLOW

The renal blood flow in man is usually measured by the clearance technique There are some restrictions necessary in connection with this method *1)* rapid changes in the renal blood flow cannot be detected since a single clearance period covers 5 to 15 min and gives only average values during this time, 2) determinations can be carried out only when urine is produced at the same time, since for the calculation of renal blood flow, it is necessary to know the amount of the clearance substance (usually paraminohippuric acid) which is excreted in the urine For example, in shock and in certain renal diseases, when no urine is produced, renal blood flow cannot be evaluated by this method *3)* Urine formation is also interrupted when the arterial blood pressure is less than 50 mm Hg, since the hydrostatic pressure in the glomerular capillaries is insufficient to produce a glomerular filtrate

In order to get a more exact picture of the renal hemodynamics, it was necessary, therefore, to turn to animal experiments Nearly all the direct and indirect methods for the measurement of blood flow described in Chapter 7 have been applied in these animal experiments However, these methods fail to elucidate the pattern of distribution of the *intrarenal* blood flow Methods have been developed separately to determine the blood flow of the renal cortex and medulla *1)* Under certain circumstances the accumulation rate of a substance in a tissue can be used as a measure of the blood flow This method has been applied

to the determination of blood flow to the different renal areas When radioactive albumin is introduced into the arterial circulation, the radioactivity builds up in the tissue The rate of accumulation of the radioactivity is a function of the blood flow to the region 2) Another method utilizes the translumination of the different parts of the kidney by implanted micro light sources

Regional blood flow rate is calculated by means of local blood volume and local circulation time

$$F\,(\text{ml}/100\text{ g}/\text{min}) = \frac{\text{vascular volume}/100\text{ g}}{\text{mean circulation time}}$$

Local vascular volume is calculated utilizing the ratio

$$\frac{\text{Hemoglobin concentration in tissue}}{\text{Hemoglobin concentration in blood}} = \text{vascular volume}/100\text{ g tissue}$$

To obtain mean circulation times of blood in the cortex and outer and inner medulla, dye-dilution curves in these areas were recorded simultaneously after a single injection of Evans blue or Cardio green into the renal artery (see Fig 3 180) Recordings were taken by small photoelectric probes placed on the cortical surface and on one side of the papilla or further up into a calyx Whereas in the cortex a photoelectric reflectometer was pierced under the capsule, the medullary photocells received light transmitted from a small tungsten bulb at the tip of a hypodermic needle pierced through the tissue Local mean circulation times were calculated from the dye-dilution curves

Thorburne and associates (1963) used the inert gas method in conscious dogs to determine intrarenal blood flow distribution After a single injection of ^{85}Kr into the renal artery, the disappearance curve of the gamma emission was monitored with a scintillation probe placed on the body surface above the kidney Assuming an equilibration of krypton between blood and tissue in a single passage through the capillary bed, and assuming an extremely small recirculation, the rate of disappearance from the kidney will depend upon the nutrient or capillary blood flow (F)

$$F(\text{ml}/100\text{ g}/\text{min}) = (k \times \lambda \times 100)/p$$

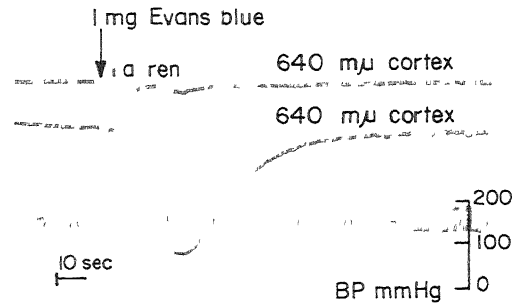


Fig 3 180 Dye dilution curves in the renal cortex and medulla of the dog after rapid injection of Evans blue into the renal artery (↓ i a ren) The passage of Evans blue was measured by small photocells which separately picked up reflected light from the cortex (*upper curve*) and transmitted light through the medulla (*middle curve*) The photocells were sensitive at the Evans blue absorption wave length (640 mμ) The lower curve represents systemic blood pressure (BP) (From Kramer et al 1960)

F = arterial inflow rate per 100 g tissue
k = slope of disappearance curve
λ = partition coefficient for the inert gas between tissue and blood
p = specific gravity of the tissue

In contrast to other organs, the decay curves obtained in the kidney are nonexponential and can be described by a series of exponentials each associated with different blood flow rates through localized regions of the kidney Four regions could be differentiated *1*) cortex, *2*) juxtamedullary cortex and outer medullary, *3*) inner medullary, and *4*) perirenal and hilar fat

NORMAL RENAL BLOOD FLOW

Under resting conditions, in the recumbent position, and with a normal mean arterial blood pressure, about 1200 ml of blood per min pass through both human kidneys, or about 20% of the cardiac output This is called the renal fraction of the cardiac output Since the average weight of both kidneys is about 300 g, normally 1 g of kidney is perfused by 4 ml of blood per min However, this value is only an average for the whole kidney since there are great differences in the distribution of the blood supply within various renal areas Blood flow values per gram of tissue have been found to be similar in dogs (4 to 5 ml /g /min) However, smaller animals have a greater renal perfusion rate Values up to 10 ml /g /min have been obtained in rats by the PAH clearance method For a better comparison of

values obtained from different laboratories, it has become the custom to express renal blood flow in milliliters per gram per minute

In the kidney, there are two capillary beds connected in series the glomerular capillaries and the peritubular capillaries They are separated by the vasa efferentia, which functionally and anatomically behave like arterioles The main resistance to flow in the kidney is divided almost equally between the vasa afferentia (preglomerular) and vasa efferentia (postglomerular), and thus the main pressure drop occurs in these segments (Fig 3 181) It has been demonstrated by micropuncture studies in the cortex of mammalian kidney that the hydrostatic pressure in the peritubular capillaries is usually 14 to 18 mm Hg, and is about the same as the hydrostatic pressure in the proximal tubules At the present time, there are no direct measurements of the hydrostatic pressure in mammalian glomerular capillaries However, by indirect calculations, this pressure has been assumed to be normally in the range of 60 to 70 mm Hg

Ever since the investigations of Wirz and associates (1951) showed the significance of the renal medulla upon the concentrating mechanism of the kidney and the functional difference of this area from the renal cortex, interest has grown in the distribution of total renal blood flow to the component parts of the kidney Up to the present, all the data on the intrarenal distribution of blood flow are the results of experiments in dogs because of the operative procedures involved in obtaining these measurements The results show that the perfusion rate in the renal cortical tissue is much greater than in the medulla (Fig 3 182), and indicate that only 1 to 2% of the total renal blood flow passes through the inner medulla This low blood flow in the inner medulla is associated with a low velocity of the blood in the vasa recta of the medulla The mean circulation time of the blood through the vasa recta is about 50 sec , and only about 2 5 sec through the cortical capillaries (Fig 3 180) On the other hand, the vascular volume in the cortex and in the medulla seems to be the same, in both areas, the vascular volume has been calculated to be about 20 to 22% of the tissue volume. More than one factor is known to cause this low velocity of blood in the vasa recta. The resistance to flow in these vessels is

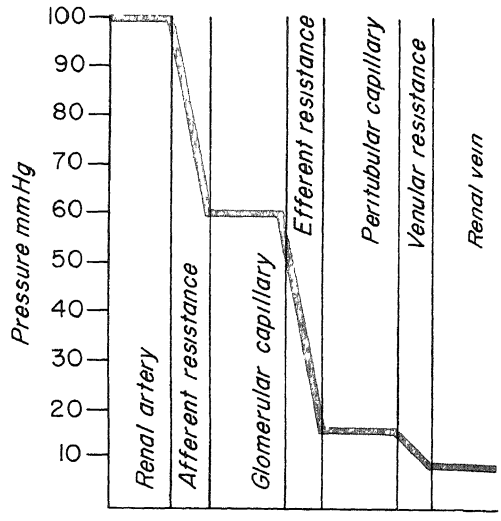


Fig 3 181 Pressure gradients in the mammalian renal circulation

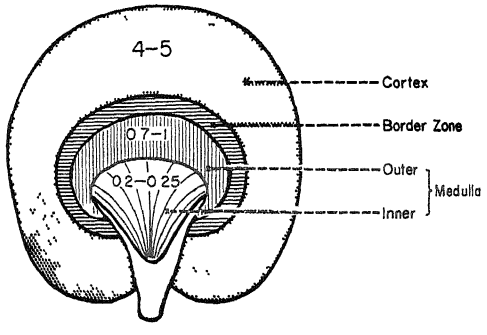


Fig 3 182 Blood flow rates (milliliters per gram per minute) in different areas of the dog kidney

very high because of the unusual length (as compared to other capillaries) of the vasa recta, which may be as long as 20 mm The composition of the blood is changed as it passes through the medulla, especially when the end urine is highly concentrated The medullary blood becomes more and more dehydrated as it reaches the tip of the papilla, by transmembrane shunting of water from the descending to the ascending limb of the vasa recta This water-shunt mechanism results in an increase of the blood viscosity Thus, the unusual length of the vasa recta and increasing blood viscosity all contribute to the low velocity of medullary blood flow Table 3 22 summarizes data on intrarenal circulation times, vascular volumes and blood flow rate

Table 3 22
Intrarenal circulation times vascular volumes and blood flow rate[a]

Source	Weight (% of Total Kidney)	Intrarenal Circulation Times[b]	Vascular Volume		Blood Flow Rates		
			ml /100 g tissue	ml /100 g kidney	ml /100 g tissue/ min	ml /100 g kidney/ min	% of total renal blood flow
		min					
Cortex	70	0 021	19 2	13 5	458	321 0	92 5
	75				472	354 0	94 1
Outer medulla	20	0 086	19 2	3 9	112	22 4	6 5
	15				132	29 0	5 3
Inner medulla	10	0 75	22 0	2 2	29	2 9	1 0
	10				17	2 0	0 6
					22		
Compartment I + II		0 095 (0 073)		17 3		209 5	96 9
Compartment III		1 4 (1 1)		6 1		6 5	3 1

[a] From Thurau, 1964
[b] Figures in parentheses represent values for erythrocytes, all other values were obtained from plasma

INTRARENAL HEMODYNAMICS AND RENAL FUNCTION

Intrarenal hemodynamics and renal function are so interrelated that change in one will always affect the other An increase in the cortical blood flow, if correlated with elevated glomerular capillary pressure, causes an increase in the glomerular filtration rate In these circumstances, the excretion of osmotically active substances will be increased This latter effect is partially caused by an increase in the tubular load, i e , a greater amount of osmotically active substances offered to the tubules by the increased glomerular filtration rate On the other hand, in the presence of an increase in cortical blood flow, there is a concomitant increase in the peritubular capillary pressure which, by itself, may affect the total excretion by means of decreased proximal tubular reabsorption

Medullary blood flow seems to be important for the concentrating ability of the kidney In the renal medulla, the osmotic pressure progressively increases from the base to the tip of the medulla and becomes a determinant in the concentration of the final urine delivered to the renal pelvis. An increase of blood flow in the medulla tends to wash out the osmotically active substances from this area, decreasing the osmotic pressure and causing a lower concentration of the final urine In this manner, renal function is affected by primary changes in medullary blood flow and, conversely, medullary blood flow may be affected by primary changes in renal function In water diuresis, the medullary blood flow is greater than in antidiuresis The cortical blood flow and the glomerular filtration rate are unchanged under these circumstances

AUTOREGULATION OF RENAL BLOOD FLOW

The blood flow (F) in an organ is directly proportional to the arteriovenous pressure gradient (Pa-v), and inversely proportional to the resistance (R), i e , $F = (Pa\text{-}v)/(R)$ In certain organs (e g , the skin or the pulmonary circulation), when only the arteriovenous pressure gradient increases, the blood flow increases more than the increase in pressure gradient, meaning that there has been a concomitant decrease in the resistance caused by the passive distention of the vessels with pressure The pattern of behavior of the renal blood flow is different, it reacts in the manner just described at perfusion pressures only up to 90

mm Hg However, when the arterial perfusion pressure rises above 90 mm Hg, the resistance increases so that the total blood flow in the kidney remains nearly constant (Fig 3 183, curve N) Above 250 mm Hg, the blood flow again becomes directly dependent on the arterial pressure This phenomenon is known as the autoregulation of renal blood flow because it is also present after cutting the renal nerves, after injecting ganglioplegic drugs, and in the isolated perfused kidney

Simultaneous measurements of blood flow in the cortex and inner medulla of the kidney have shown that autoregulation is present solely in the cortex Since only about 1 to 2% of the total renal blood flow passes through the inner medulla, a solitary increase in medullary blood flow does not actually cause a measurable increase of total renal blood flow Therefore, in the presence of a rising arterial perfusion pressure, the total renal blood flow

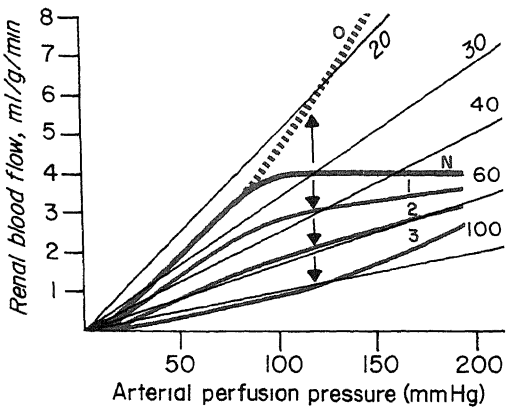


Fig 3 183 Pressure-flow curves in the kidney The thin lines through the origin of the graph represent equal resistances to flow expressed in terms of

$$\frac{\text{mm Hg}}{\text{ml}}$$

per gram per minute

Curve N pressure flow curve under normal conditions The increase of resistance above 80 mm Hg is independent of extrinsic factors (see autoregulation of renal blood flow) *Curves 1 to 3* pressure flow curves with increasing vascular tone secondary to various extrinsic factors such as increase sympathetic stimulation or epinephrine With increasing extrinsically induced vascular tone the pressure flow curves become more and more linear simultaneous to a decrease in autoregulation *Curve 0* pressure flow curve after paralyzing the smooth muscles of the renal vessels Note the abolition of autoregulation also note that under normal conditions (*curve N*), there is no myogenic tone in the renal vessels between 0 to 80 mm Hg

remains nearly constant in spite of an increasing medullary blood flow Hence, it is more accurate to limit discussion of autoregulation soley to the cortical blood flow

There are at the moment, four theories which attempt to explain this special behavior of the renal blood flow

Myogenic Theory When the arterial perfusion pressure is increased, the vessels become distended The smooth muscles, especially in the preglomerular vessels, react to this distention by shortening, thereby constricting the vessels so that a new high resistance is established This implies that the tonus of the vessels is increased by an autonomic myogenic reaction to increasing arterial perfusion pressure [1] It has been demonstrated in the dog kidney that when the arterial perfusion pressure is suddenly increased, the blood flow is only increased passively during the first 3 to 4 sec Following this, the resistance in the vascular bed increases to such an extent that the blood flow nearly ceases at about 15 sec After 20 to 30 sec the blood flow increases to the normal level of the autoregulation (Fig 3 184) This initial latency and the phasic changes of the resistance are in accord with the contracting properties of smooth muscle The autoregulation can be abolished by drugs which paralyze the smooth muscles (potassium cyanide, papaverine, and procaine (novocaine) in high concentration) Likewise, the phasic changes of the resistance, caused by rapid pressure increases, are removed by these drugs (Fig 3 184)

There are experimental indications that the autoregulatory mechanism resides within the preglomerular vessels The glomerular filtration rate is controlled, among other things, by the intraglomerular capillary pressure The constancy of the glomerular filtration rate and its independence from arterial perfusion pressure changes above 90 mm Hg, constitutes

[1] Contraction of the smooth muscles as the result of elevated intravascular pressure has been described for many vascular systems In contrast to the kidney, these myogenic reactions are not powerful enough to constrict the vessel to such an extent that blood flow remains constant in spite of an elevated perfusion pressure It is not necessarily true that the single muscle cell of the renal vessel reacts differently than in other organs A spatial arrangement of the muscle cells, specially oriented, can increase the net effect of the single cell contraction on the diameter of the vessel

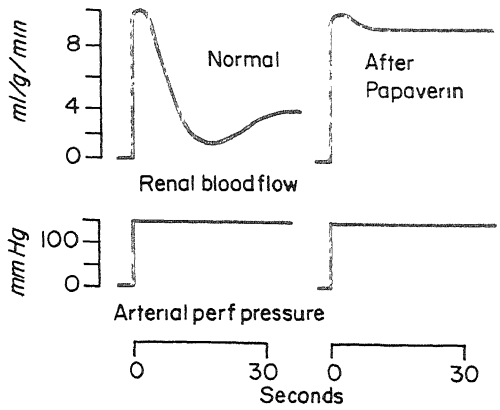


Fig 3 184 Pattern of flow in the renal artery of dogs following sudden increase in the arterial perfusion pressure from 0 to 150 mm Hg before and after vascular muscle paralysis by papaverine The phasic changes in renal blood flow in the normal kidney indicating a concomitant inverse change in resistance can be abolished by papaverine The steady state blood flow of 4 ml/g/min at 150 mm Hg in the control (autoregulated kidney) is increased under papaverine to about 9 ml/g/min (see also Figure 3 183 difference between curve N and curve 0 at 150 mm Hg) (Modified after Thurau 1959)

indirect evidence of a constant, intraglomerular capillary pressure Furthermore, by paralyzing the vascular smooth muscles with papaverine, the glomerular filtration rate now becomes directly proportional to the arterial perfusion pressure Thus, there is strong evidence to support the contention that the autoregulatory mechanism depends upon the integrity and reactivity of the preglomerular vascular muscles

Cell Separation Theory. In a vascular system in which small vessels branch off from a large one, the small vessels receive cell-poor plasma from the outer zone of blood in the large vessel, because the erythrocytes remain in the axial stream of the large vessel This separation of plasma and erythrocytes at the bifurcation of vessels is called plasma skimming or cell separation For the kidney, it is assumed that this cell separation takes place to an unusual degree at the origin of the afferent arterioles The blood remaining in the interlobular arteries becomes more and more concentrated with erythrocytes and flows through special erythrocyte-shunts directly into the veins without flowing through the peritubular capillaries, which are perfused only by cell-poor plasma The degree of the cell separation is supposed to be dependent on the arterial perfusion pressure so that blood with a hematocrit up to about 80% in the distal part of the interlobular arteries can be produced In this range of hematocrit the viscosity of the blood is high and varies greatly with small changes in hematocrit This variable viscosity which is dependent on the arterial blood pressure, is believed to be the cause of the variable resistance responsible for the autoregulation According to this theory, a cell-free perfusion of the kidney should not show any autoregulation However, it has been demonstrated by different investigators that autoregulation does occur even when the kidney is perfused by dextran, a cell-free plasma, or blood with a very low hematocrit It would furthermore, be expected that the oxygen saturation of the blood in the distal peritubular capillaries should be lower than that in the mixed venous blood However, investigations have shown that the oxygen saturation in this area is not different from that in mixed venous blood Also, the passage times of plasma and erythrocytes from the renal artery to the renal vein are the same, but, according to the cell separation theory, they should differ

Intrarenal Pressure Theory. By insertion of a 24 gauge needle into the renal tissue, it was found that the intrarenal tissue pressure measured in this way increases disproportionately with increasing arterial perfusion pressure From this data, it is calculated that the elevated tissue pressure has a clamping effect on vascular geometry and resistance to flow This assumes that the needle pressure selectively measures only normal tissue pressure and that no vessels have been ruptured by the needle insertion In any case, measurements of the intracapillary and intratubular pressures in the rat kidney by the micropuncture technique have shown that these pressures do not change when the arterial perfusion pressure is elevated above 90 mm Hg

Intrarenal Sodium Feedback Theory at the Level of the Juxtaglomerular Apparatus. The single nephron unit in the mammalian kidney is arranged in such a way that the end of the ascending limb of Henle's loop comes in direct contact with the vascular pole of its own glomerulum At the point of contact the renin-containing cells are located in the wall of the afferent arteriole The tubular cells, at the point of connection are called *macula densa cells* and are in contact at their luminal

side with tubular fluid, leaving the ascending limb of Henle's loop, and at the peritubular side with the granular cells of the afferent arteriole Goormaghtigh (1939) suggested the macula densa cells were a sensing device by which changes in concentration of the tubular fluid could regulate the renin release in the afferent arteriole Since renin is the enzyme necessary for angiotensin formation, preglomerular vasoactions could be initiated by such a mechanism Recently it was shown by Thurau and Schnermann (1965) that an increase in sodium concentration in the tubular fluid at the macular densa cells leads to a reduction in glomerular filtration rate Assuming the kidney showed no autoregulation, then the increase in arterial blood pressure would not only increase renal blood flow but also glomerular filtration rate and tubular sodium load In order not to lose the excess filtered sodium, the metabolically linked tubular sodium reabsorption in the proximal tubule would have to increase instantaneously Even if equal fractions of filtered sodium are reabsorbed in the proximal tubule, a greater amount of sodium during increased GFR would be delivered in the more distal tubular segments Sodium concentration at the macular densa cells would therefore increase and reflect the increased GFR and RBF The increased sodium concentration at the macula densa cells lowers the glomerular filtration rate by preglomerular formation of angiotensin which leads to a preglomerular vasoconstriction and concomitantly to a fall in filtration rate On the basis of this mechanism, renal autoregulation would primarily be a phenomenon which keeps sodium load as the major determinant for renal metabolism constant, and autoregulation of blood flow under normal conditions would be a consequence of such a regulation

NERVOUS CONTROL OF RENAL BLOOD FLOW

The highly developed plexus renalis contains sympathetic and parasympathetic fibers However, under normal physiological conditions, not too much is known of the significance of this rich nerve supply to the kidney The interpretation of the results of experiments with regard to the nervous control of renal blood flow is complicated by the fact that the vascular system of the kidney has an autonomous and variable tonus which is independent of any innervation (see autoregulation of renal blood flow) and may interfere with nervous vascular reactions For a better understanding of this vascular behavior, see Figure 3 183, which shows pressure flow curves resulting either from change of nervous vascular tonus or from autonomous tonus

Much investigation has been going on to gain information about the nervous control of the renal blood flow under *resting conditions* Renal blood flow has been measured before and after denervation of the kidney

In *unanesthetized* man and animals in recumbency, the renal blood flow of the denervated kidney does not differ from that of the innervated kidney This implies that under resting conditions there is no sympathetic nervous influence in the renal vascular bed Furthermore, the nonexistence of a sympathetic tone of the renal vascular bed, under resting conditions, has been confirmed by animal experiments in which depressor reflexes in the systemic circulation were produced Stimulation of the reflexogenic areas of the carotid sinus and aortic arch causes vasodilation in the kidney only when a sympathetic tonus of the renal vessel has been *previously* produced by asphyxia or by anesthesia

In anesthetized animals, it has been shown that the renal blood flow may increase after denervation, but, since the sympathetic tone in the peripheral vascular bed is usually augmented during anesthesia, the denervation, therefore, eliminates only the unphysiological sympathetic stimulation produced by the anesthesia, and thereby permits return of the blood flow to normal values Accordingly, in a dog where one kidney is denervated, the renal blood flow decreases only in the innervated kidney during pentobarbital injection The blood flow of the denervated kidney shows no change

Neurogenic vasoconstriction of the renal vessels is produced by severe hypoxia, i e , arterial oxygen saturation below 50% However, in the range of decreasing arterial oxygen saturation between 100 and about 50%, there seems to be no change in renal blood flow, and this is true either in anesthetized dogs or in unanesthetized man The vasoconstrictor effect of severe hypoxia can be eliminated in the dog by abolishing the chemoreceptors in the carotid sinus and aortic arch, indicating that the constrictor response to hypoxia is induced by stimulating the peripheral chemoreceptors.

Vasconstriction in the dog kidney also occurs during hypercapnia while breathing a mixture of 5 to 30% CO_2 in oxygen In these experiments, the pH of the blood simultaneously decreased from 7 49 to 7 21 Since the constriction is abolished by renal denervation hypercapnia and this degree of acidosis seem to produce only neurogenic rather than local vasconstriction

A decreased discharge rate of the pressoreceptors in the carotid sinus or aortic arch in dogs, caused by lowering the intrasinusal or intra-aortic blood pressure, or by damping the pulse pressure in these areas, leads to neurogenic vasoconstriction in the kidney as well as in other organs

In connection with the nervous control of the renal blood flow, the question also arises as to whether or not there exists a nervous control of the function of the tubular cells After denervation of the kidney, an increase of sodium excretion is often found, suggesting a decreased absorptive capacity of the tubular cells However, in nearly all these experiments, the glomerular filtration rate simultaneously increased It is well known that this increase of the glomerular filtration rate can cause an elevated sodium excretion without altered tubular function Moreover, those experiments in which the excretion of electrolytes was increased *without* a corresponding increase in the glomerular filtration rate, by no means prove a primary changed tubular function after denervation As yet, it is still unknown whether changes of the intrarenal distribution of blood flow occur after denervation since, for example, even small changes in medullary blood flow can affect the composition of the end urine

EXERCISE, POSTURE AND CENTRAL BLOOD VOLUME EFFECTS ON RENAL BLOOD FLOW

In the upright exercising man, renal blood flow is diminished when compared to blood flow under resting conditions while supine If, however, exercise is performed in recumbency, the blood flow in the kidney remains nearly unaffected Also, the dog, which usually does not change its posture during exercise, shows no change in renal blood flow when running 5 6 to 10 m.p h at a 15° slope on a treadmill

These findings give rise to the supposition that the upright position per se, in which exercise is usually performed leads to the decrease in renal blood flow Indeed it has been found that the usual response on changing from the recumbent to the standing position is a decrease in renal blood flow Under the extreme conditions of prolonged standing, the renal blood flow may decrease by about 50% of the control value These changes are accompanied by similar reductions in glomerular filtration rate and urine output Since the arterial blood pressure is changed little or is unchanged, the reduction in renal blood flow is primarily caused by an elevated resistance to flow This is only part of a generalized elevation in total peripheral resistance secondary to the upright position This enhanced peripheral resistance is caused by an augmented neural vasoconstriction rather than by a humoral mechanism The afferent mechanisms involved are poorly understood, but it is believed that this can be effected through reduction of central blood volume The central blood volume in the thorax can be reduced to subnormal values by means of positive pressure breathing It has been shown, that under these circumstances, the renal resistance to flow increases Since the blood volume in the chest is also diminished in the upright position, this reduction in central blood volume, induced by changing from the recumbent to the standing position, may be one factor involved in the renal constrictor mechanism Conversely, an increase in central blood volume above normal by blood infusion, tilting of the body to the head down position, or negative pressure breathing, should produce a neurogenic vasodilation in the kidney But, as pointed out in the preceding section on nervous control of renal blood flow, under normal resting conditions there are *no* sympathetic discharges to the renal vessel, and hence a neurogenic vasodilation is nonexistent It is not surprising, therefore, that an increase in central blood volume fails to result in a renal vasodilation However, there is another interesting effect of an increased central blood volume upon the kidney which should be briefly mentioned here It has been found that elevation of central blood volume simultaneously increases the urine output Since renal blood flow and also the glomerular filtration rate are not altered under these conditions, and, since the resultant diuresis is caused primarily by an increased free

water excretion (i e , without concomitant excretion of osmotically active substances), there is strong evidence that an increase in central blood volume inhibits antidiuretic hormone (ADH) secretion of the neurohypophysis It is suggested that this ADH-inhibition is reflexly induced by stimulation of stretch receptors located predominantly in the left atrial wall and sensitive to an increase in central blood volume (see also p 3-186)

Returning to the discussion renal blood flow during exercise in man, there is no doubt that the change to the upright position per se is one factor involved in the decreased renal blood flow But, reductions in renal blood flow down to 20% of the control value, as measured in severe exercise in man, can hardly be explained by this orthostatic mechanism alone It is a matter for further investigations to establish the other factors which determine the changes in renal blood flow during exercise

EFFECT OF SOME PHARMACOLOGICAL AGENTS UPON RENAL CIRCULATION

Epinephrine and *norepinephrine* both have powerful vasoconstrictor effects upon the renal vessels Each drug acts upon smooth muscles, and its action can be abolished by adrenergic blocking agents Single intravenous injection of 2 μg /kg of epinephrine causes complete cessation of renal blood flow During continuous intravenous infusion of low doses (10 to 15 μg /min) of epinephrine or norepinephrine in man, there occurs a 30% decrease in renal flow without change in the glomerular filtration rate The result indicates that low concentrations of the pressor agents predominantly produce efferent arteriolar constriction Since the systemic arterial pressure is concomitantly elevated, a small rise in resistance must also occur in the afferent arterioles or, otherwise, there would occur an increase in the glomerular filtration rate Similar results have been found in the denervated kidney in man The effect of 2 to 10 μg /min of l-norepinephrine, given intravenously, has been studied in the homotransplanted kidney of an identical twin recipient and in the innervated kidney of the donor In both kidneys the renal blood flow decreased 25%, when given 2 μg l-norepinephrine, and 50%, when given 10 μg , without changes in the glomerular filtration rate, revealing the similarity in the response of the innervated and denervated kidney to l-norepinephrine in unanesthetized man Despite the decrease in renal blood flow following minute doses of epinephrine, there occurs a moderate swelling of the kidney Experiments in dogs have shown a simultaneous increase in renal blood volume It has, therefore, been concluded that an increase in resistance to flow also occurs in postcapillary segments (probably venular) which leads to an elevated intracapillary pressure and a resultant distention of the renal capillaries

Pitressin is an extract of the posterior pituitaries of oxen and hogs with an antidiuretic-vasopressor action Infusions of 1 to 2 mμ/kg /hr, with which significant antidiuretic effects can be produced, have no effect upon the renal blood flow in man Higher, nonphysiological doses increase the resistance to flow in the kidney by direct action on the renal vessels Repeated injections elicit tachyphylaxis

The injection of *serotonin* (5-hydroxytryptamine) into the renal artery in dogs, produces constriction of the renal vessels This constrictor effect has been shown in innervated as well as in denervated kidney and persists after the application of sympatholytic agents Thus, serotonin probably acts directly on the smooth muscles However, *intravenously* injected serotonin may be followed by a renal vasodilation reflexly induced by the action of this drug upon depressor receptors whose afferent fibers arise within the chest and are mediated through the vagus nerve In patients with malignant carcinoid, the urinary excretion of serotonin or its metabolic products is elevated secondary to the high circulating blood level In these patients, the glomerular filtration rate and renal blood flow are usually diminished

1-Hydrazinophthalazine (Apresoline) produces a vasodilation in the kidney which persists for hours following a single injection The mechanism of the vasodilatory effect is still unknown It seems that Apresoline has no direct paralyzing effect upon the renal vessels since injection into the renal artery produces an initial vasoconstriction, followed in several minutes by the onset of vasodilation, the latter becoming more marked during the next 30 min and persisting for 2 to 3 hr

REACTIVE HYPEREMIA

In both cardiac and skeletal muscles, and in some other organs, the blood flow temporarily increases over the control value after releasing an arterial clamp This overshoot of blood flow, lasting from several seconds to several

minutes, depends upon the duration of occlusion and may exceed the estimated blood flow debt incurred during the occlusion period If the term reactive hyperemia is restricted to this overshoot, then by definition, the renal circulation does not show this phenomenon After releasing an occlusion of the renal artery there occurs a brief 3- to 5-sec overshoot of blood flow (Fig 3 185), which however is independent of the duration of occlusion and is, in no instance, greater than the estimated blood flow debt Furthermore, after interruption of the renal blood flow for several minutes, the re-established blood flow, following the initial overshoot, usually decreases temporarily to subnormal values

HYPOTHERMIA AND RENAL BLOOD FLOW

At the onset of the cooling phase, as the body attempts to maintain a normal temperature (shivering phase), the renal blood flow increases 20 to 45% above the control value This has been observed utilizing para-aminohippurate (PAH) clearances in both anesthetized and unanesthetized man and animals This initial increase in renal blood flow can be abolished by inhibiting shivering with d-tubocurarine As the rectal temperature falls below 32° C (90° F), the renal blood flow progressively decreases and, at 27° C (80° F) is only 50% of the normal value This decrease in blood flow is also observed in the isolated dog kidney which was perfused by cooled blood, and in which the renal blood flow was directly measured A major part of the decreased blood flow is caused by an increased resistance related to the augmented viscosity of cooled blood, but this factor alone, however, is insufficient to explain the degree of change in blood flow It is still unknown which other nervous or humoral factors are involved

Regional cooling such as the immersion of an extremity into ice water for some minutes, sufficient to produce a painful sensation, in man is attended by a 20% decrease in renal blood flow This is probably caused by a nervous reflex, stimulating sympathetic discharges to the renal vessels

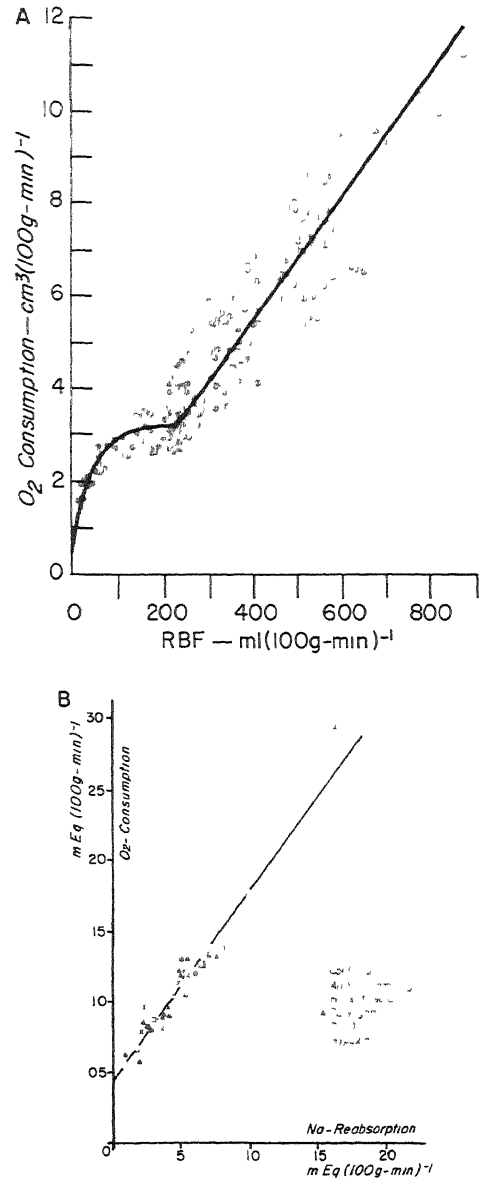


Fig 3 185 Correlation of renal oxygen consumption with renal blood flow (A) and tubular sodium reabsorption (B) Note the exponential increase of oxygen consumption at low renal blood flow values Sodium reabsorption was varied by decreasing glomerular filtration rate (aortic clamp) hypertonic sodium chloride infusion saluretic drugs and hypoxia (From Kramer and Deetjen 1961)

ARTERIOVENOUS OXYGEN DIFFERENCE AND OXYGEN UPTAKE OF THE KIDNEY

The oxygen difference between arterial and venous blood in the human kidney (AV oxygen difference) is about 1 7 ml /100 ml of blood, which is small compared with an AV oxygen difference of 4 to 5 ml /100 ml of blood between arterial blood and mixed venous blood of the right heart From the renal AV oxygen difference and the total renal blood flow (1200

ml /min), the oxygen uptake of both human kidneys is calculated to be about 21 ml /min In isolated dog kidney, as well as in dogs' kidneys in situ, the AV oxygen difference has been found to be fairly constant in spite of renal blood flow changes, which means that the oxygen consumption in the kidney parallels the blood flow

The observation of Zerahn (1956) that the oxygen uptake of an isolated frog skin above a resting level is proportional to active sodium transport, paved the way for an explanation of the direct relation between RBF and oxygen consumption It is well established that renal sodium reabsorption is an active process and represents the overwhelming fraction of all tubular processes Since under normal conditions, GFR varies with RBF, tubular sodium reabsorption is high at high RBF and accounts for the elevated oxygen consumption (see Fig 3 185) The constant AV difference of oxygen only above 60 mm Hg BP is consistent with the onset of glomerular filtration at this range of BP, whereas below this pressure the AV difference of oxygen is inversely correlated to RBF as in most other organs It is clear that a direct correlation between RBF and oxygen consumption is only obtained as long as GFR changes with RBF In osmotic diuresis and during ureteral occlusion oxygen consumption is lowered at an increased RBF since GFR is diminished

The AV oxygen difference in the renal medulla of anesthetized dogs has also been indirectly calculated by means of the oxygen tension of the final urine, which is assumed to be in equilibrium with the medullary venous blood Usually, the oxygen tension in the urine is less than in the mixed renal venous blood and in these experiments, the difference during air breathing was 25 ± 3 mm Hg Thus, by this method, the AV oxygen difference in the medulla has been calculated to be 8 5 ml /100 ml of blood, which would indeed be very high However, there are indications that oxygen pressure in the medulla may also be influenced by the counter-current exchange mechanism In this case, oxygen pressure would progressively decrease to the tip of the medulla by transmembrane shunting of oxygen from the descending to the ascending limb of the vasa recta The low oxygen tension of the final urine would represent only the low oxygen pressure at the tip of the medulla rather than the oxygen pressure of the medullary venous blood and, therefore, could not be used for calculations of the medullary AV oxygen difference

CONCLUSIONS

Renal hemodynamics exhibit certain peculiarities which need re-emphasis Under basal conditions, there exists no neurogenic tone in the renal vessels, and the kidney is protected against elevations of arterial blood pressure by an autonomous mechanism which usually maintains the cortical blood flow, the glomerular filtration rate and the intrarenal capillary pressure at constant levels A neurogenic vasodilation is generally nonexistent and only vasoconstriction can be produced via the nervous system Renal hemodynamics are poorly understood solely by using measurements of *total* renal blood flow The renal artery supplies blood at varying flow rates to the differently functioning areas, and function and blood flow seem to be mutually interdependent, this has only been demonstrated in animals The human kidney, however, still holds secret its distribution of intrarenal blood flow

Fetal and Neonatal Circulation

In Figure 3 186, diagrams of the fetal, transitional (neonatal) and adult circulations are shown In the fetus, the left and right heart halves act essentially as parallel pumps, discharging blood into the systemic and placental circuits Blood from the right heart largely bypasses the lung circulation by flowing through the *ductus arteriosus* into the aorta In contrast, the two sides of the heart in the adult pump chiefly in series, blood passes sequentially through the right heart and pulmonary circulation to the left heart and systemic circulation It is only after birth that a large pressure differential develops between the pulmonary and systemic arterial systems Consequent to this the relative work load (pressure × volume) of the left heart is increased greatly over that of the right This difference is reflected in the relative masses of the adult left and right ventricles (about 2 1) The transition from fetal to adult circulatory patterns is accomplished by functional and anatomical closure of fetal flow paths (*ductus arteriosus, foramen ovale, ductus venosus*), expansion of the pulmonary circuit with expansion of the lungs,

and interruption of the placental circulation Failure of *ductus arteriosus* or *foramen ovale* closure are frequent congenital cardiac defects with hemodynamic consequences proportional to their degree of ultimate patency Functional closure of these fetal bypass routes is almost complete in the early neonatal period, but anatomical closure requires weeks or months (Fig 3 187)

CIRCULATION IN THE FETUS

The placenta serves as the "fetal lung" and access port to the maternal circulation for uptake of nutritional substances and excretion of metabolic wastes The major circulatory circuitry of the fetus is diagrammed in Figure 3 186 (*upper*)

Most studies on the physiology of the fetal and neonatal circulation have been carried out on sheep and goats These species differ from man in having two umbilical veins rather than

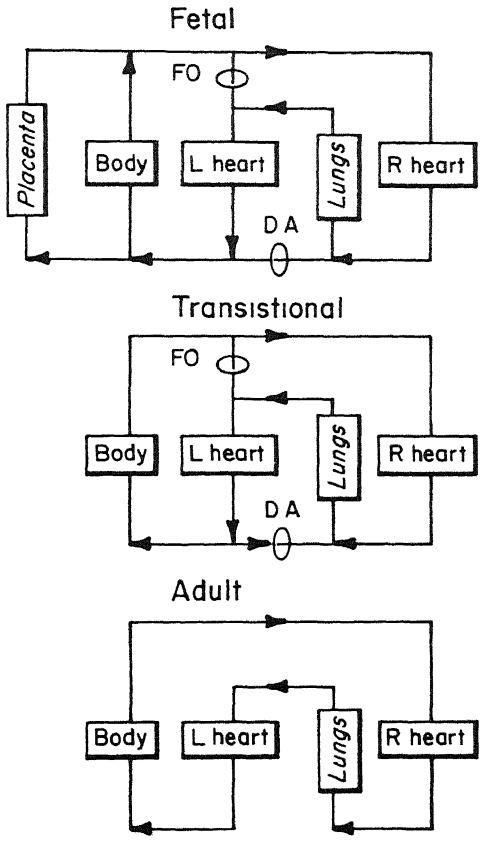


Fig 3 186 The fetal transitional (neonatal) and adult types of circulation (Redrawn from Dawes, 1968)

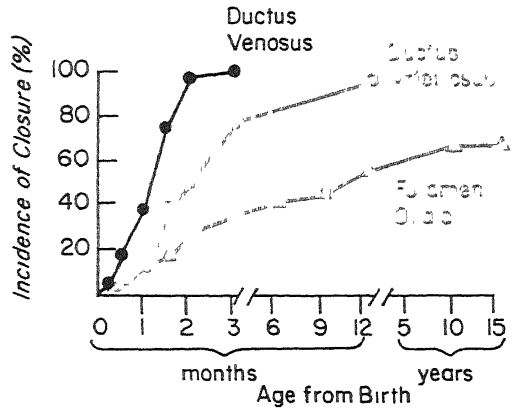


Fig 3 187 Anatomical closure of the special vascular channels of the fetus after birth (Redrawn from Dawes 1968)

one They possess cotyledonary placentas and have a relatively smaller brain Results and interpretations reported in the literature contain unresolved differences in certain details (Dawes, 1968, Assali, 1968)

In the fetus, the combined outputs of the left and right ventricle are about equally divided between the placenta and fetal membranes and the body of the fetus itself These two circulations are arranged in parallel (Fig 3 186 *upper*) and the following percentage distributions of flow have been calculated from data on lambs (Dawes, 1968)

Placenta and fetal membranes	57%
Fetal lungs	10%
Foreparts of the body	15%
Hindparts of the body	18%

The blood circulating through the placenta, enters with an oxygen saturation of about 58% and returns in the umbilical veins 80% saturated (contrasted with 98% oxygen saturation of maternal arterial blood) The fetal *ductus venosus* carries some of this blood directly into the inferior vena cava, bypassing the liver Blood from the lower trunk and extremities (26% saturated) and from the liver flow together with the umbilical vein blood in the inferior vena cava.

Anatomically, the foramen ovale opens directly off the inferior vena cava so as to favor flow of blood into the left atrium There is evidence that the edge of the interarterial septum, the *crista dividens*, divides the inferior vena cava blood into two unequal streams According to this view the larger stream is

composed mainly of umbilical vein blood and is shunted through the *foramen ovale* into the left atrium The smaller stream mixes in the right atrium with superior vena cava blood from the upper body and with blood from the heart This less saturated blood largely passes into the right ventricle and is pumped out the pulmonary artery Pulmonary arterial resistance is relatively high, the pressure about 5 mm Hg above that in the aorta and, consequently, more than two-thirds of right ventricular output flows through the *ductus arteriosus* into the aorta This blood enters the aorta at a point distal to the origins of the arteries to the head and upper limbs, and is thus directed to the posterior body and umbilical arteries The larger stream of highly oxygenated blood from the *foramen ovale* and blood entering the left atrium from the pulmonary veins mix and are pumped out of the left ventricle into the ascending aorta This blood reaches the arteries supplying the upper extremities and head before mixing with the highly desaturated blood from the *ductus arteriosus* and about one-third of it passes to the upper body Insofar as this separation of inferior and superior caval flows takes place, the greater will be the percentage of oxygenation of the blood reaching the upper body and brain However, a difference in percentage oxygen saturation of carotid and descending aortic blood is not always found experimentally and fetuses with cardiac malformations preventing such separation, survive to term

Fetal hemoglobin has a greater affinity for oxygen than adult hemoglobin so that at equal oxygen tension fetal blood carries more oxygen than maternal blood Also fetal tissues are more resistant to hypoxia than are adult These two factors make possible adequate oxygen supply to fetal tissues, despite a relatively low oxygen tension in blood leaving the placenta

CHANGES AT BIRTH

With separation from the placental circulation, the peripheral resistance rises and asphyxia develops in the infant The pressure in the aorta rises and gasping respiratory movements expand the lungs Constriction of the muscular umbilical veins squeezes up to 100 ml of blood into the fetal veins ("placental transfusion") and ventilatory movements aid venous return. With lung expansion, pulmonary blood flow increases as pulmonary vascular resistance falls below one-fifth the *in utero* value Pulmonary venous return to the left atrium increases left atrial pressure over right atrial pressure and the valve of the *foramen ovale* is closed In response to the increased blood oxygen tension in blood flowing through the *ductus arteriosus* the smooth muscle constricts and the *ductus* is narrowed Aortic pressure now exceeds pulmonic artery pressure and flow through the *ductus arteriosus* is reversed This is the transitional (neonatal) state of the circulatory system Functional closure of the ductus is complete within 1 or 2 days after birth Anatomical closure of the *foramen ovale* and *ductus arteriosus* may take several weeks (Fig 3 187) However, normally there is little flow through these channels after the first few days With complete closure the adult separation of pulmonary and systemic vascular systems is accomplished (Fig 3 186, *lower*) with the two halves of the heart pumping in series The only shunt remaining in the adult is that through the bronchial artery and capillary system from which a small amount of blood returns to the left heart via pulmonary veins, without having gone through the right heart

SUMMARY

Fetal circulation is especially arranged to meet the needs of a rapidly growing organism living in a state of relative hypoxia These requirements include high cardiac output, partial bypass of the lung circulation, and high placental blood flow At birth, the fetus abruptly becomes an air-breather, hypoxia disappears, blood-gas exchange occurs in the lungs and the large placental flow is abolished As a result cardiac output is reduced, pulmonary artery pressure drastically declines, and the two sides of the heart function serially, rather than in parallel Blood oxygen tension is the chief determinant of the pattern of fetal blood flow distribution and the changes in this distribution which occur at birth

The *ductus arteriosus* permits up to 70% of the right ventricular output to enter the aorta and thus provides a major source of blood which flows to the posterior fetal body and placenta Because of the parallel arrangement of left and right ventricles in the fetus, both contribute to the high systemic flow required for the fetus and placenta

Blood flow to the lungs is minimal because

of the differential effect of low oxygen tension on lung vessels and *ductus arteriosus* Hypoxemia dilates the latter and constricts the former, thus favoring flow away from the collapsed fetal lungs where peripheral resistance is high and toward the aorta (and placenta) where peripheral resistance is low At birth the reciprocal effect of increased oxygen tension on *ductus arteriosus* and lung vessels results in constriction of the former and dilatation of the latter, pulmonary vascular resistance also decreases owing to expansion of the lungs with the onset of breathing Interruption of the placental circulation increases peripheral resistance in the systemic circuit Thus, pulmonary peripheral resistance and arterial pressure decrease, arterial peripheral resistance and pressure increase, shunting of blood from pulmonary artery to aorta through the ductus arteriosus is curtailed, reverses and finally ceases and cardiac output decreases

The primary function of the *foramen ovale* is probably not as was once thought to shunt highly oxygenated blood to the left ventricle aorta and head but rather to supply blood to the left ventricle, insuring that it works mechanically and develops adequately to accomodate the abrupt increase in work load at birth

The function of the *ductus venosus* is uncertain and this structure is absent or disappears early in fetal life (horse, pig) in certain species The concept that it supplies a route for direct return of highly oxygenated umibilical venous blood unmixed with fetal venous blood, is not supported by recent studies (Assali, et al 1968)

BIBLIOGRAPHY

ABEL F L AND Q R MURPHY Mesenteric renal and iliac vascular resistance in dogs after hemorrhage *Amer J Physiol* 202 978-980 1962

ADOLPH E F *Origin of Physiological Regulations* New York Academic Press 1968

ALEXANDER R S Transformation of the arterial pulse wave between the aortic arch and the femoral artery *Amer J Physiol* 158 287-293 1949

ALEXANDER R S The influence of constrictor drugs on the distensibility of the splanchnic venous system analyzed on the basis of an aortic model *Circ Res* 2 140-147 1954a

ALEXANDER R S The participation of the venomotor system in pressor reflexes *Circ Res* 2 405-409 1954b

ANDRUS E C E P CARTER AND H A WHEELER Refractory period of normally beating dog s auricle *J Exp Med* 51 357-367 1930

ANREP G On part played by suprarenals in normal vascular reactions of body *J Physiol (London)* 45 307-317 1912

ASSALI, N S (Ed) *Biology of Gestation* New York Academic Press 1968 vol 2

ASSALI N S G A BEKEY, AND L W MORRISON Fetal and neonatal circulation In *Biology of Gestation* edited by N S Assali New York Academic Press 1968 vol 2 p 51-142

AVIADO D M *The Lung Circulation* New York Pergamon 1965, vol 1 and 2

BADEER H S Contractile tension in the myocardium *Amer Heart J* 66 432-434 1963

BADEER, H S Metabolic basis of cardiac hypertrophy *Prog Cardiovasc Dis* 11 53-63 1968

BAINBRIDGE F A On some cardiac reflexes *J Physiol (London)* 48 332-340 1914

BARGMAN W Bau des Herzens In *Das Herz des Menschen* edited by W Bargman and W Doerr Stuttgart Thieme 1963 vol 1

BARKER, J M *The Unipolar Electrocardiogram A Clinical Interpretation* New York Appleton 1952

BAROLDI G The coronary arterial collateral circulation in man *Acta Cardiol* (Suppl 13) 17-37 1969

BAUEREISEN E Herz In *Kurzgefasstes Lehrbuch der Physiologie* edited by W D Keidel Stuttgart Thieme, 1970 71-99

BAUST W AND H HEINEMANN The role of the baroreceptors and of blood pressure in the regulation of sleep and wakefulness *Exp Brain Res* 3 12-24 1967

BAYLISS L E *Living Control Systems* San Francisco W H Freeman, 1966

BAYLISS W M On the physiology of the depressor nerve *J Physiol (London)* 14 303-325 1893

BEARN, A G, B BILLING AND S SHERLOCK The effect of adrenaline and noradrenaline on hepatic blood flow and splanchnic carbohydrate metabolism in man *J Physiol (London)* 115 430-441 1952

BENCHIMOL A Significance of the contribution of atrial systole to cardiac function in man *Amer J Cardiol* 23 565-571 1969

BERNE R M AND M N LEVY *Cardiovascular Physiology* St Louis Mosby 1967

BEZNAK M J Regression of cardiac hypertrophy of various origin *Canad J Physiol Pharmacol* 47 579-586 1969

BING R J Myocardial metabolism *Circulation* 12 635 647 1955

BLINKS J R AND J KOCH WESER Analysis of the effects of changes in rate and rhythm upon myocardial contractility *J Pharmacol Exp Ther* 134 373-389 1961

BLIX M Die Lange und die Spannung des Muskels *Skand Arch Physiol* 3 295-318 1891

BLUM ROBERT L, H ALPERN H JAFFE T W LANG AND E CORDAY Determination of interarterial coronary anastomosis by radioactive sperules Effect of coronary occlusion and hypoxemia *Amer Heart J* 79 244-249 1970

BOINEAU J P AND E N MOORE Evidence for propagation of activation across an accessory atrioventricular connection in types A and B pre excitation *Circulation* 41 375 398 1970

BOWDITCH H P Ueber die Eigenthumlichkeiten der Reizbarkeit welche die Muskelfasern des Herzens zeigen *Ber Math Phys Sachs Gesellsch Wissensch* (Leipzig) 23 652 1871

BOZLER E The initiation of impulses in cardiac muscle *Amer J Physiol* 138 273-282 1943

BRADLEY S E Methods for the evaluation of the splanchnic circulation *Circulation* (Proceedings of the Harvey Tercentenary Congress) Oxford Blackwell 1958 p 355

BRADLEY, S E F J INGELFINGER AND G P BRADLEY De terminants of hepatic hemodynamics In *Visceral Circulation*, A Ciba Foundation Symposium London Churchill, 1952

BRADLEY S E F J INGELFINGER G P BRADLEY AND J J CURRY The estimation of hepatic blood flow in man *J Clin Invest* 24 890-897 1945

BRADLEY S E, P A MARKS, P C REYNELL AND J MELTZER The circulating splanchnic blood volume in dog and man *Trans Assoc Amer Physicians* 66 294-302 1953

BRAMS W A, AND L N KATZ Nature of experimental flutter and fibrillation of the heart *Amer Heart J* 7 249-261 1931

BRAMWELL J C, AND A V HILL The velocity of the pulse wave in man *Proc Roy Soc (London)* [*Biol*] *Ser B* 93 298-306 1922

BRAUNWALD, E J ROSS JR, AND E H SONNENBLICK Mechanisms of contraction of the normal and failing heart *N Eng J Med* 277 794-799, 1967

BRAUNWALD, E J ROSS JR AND E H SONNENBLICK

Mechanisms of Contraction of the Normal and Failing Heart Boston Little, Brown 1967, 1968

BRECHER G *Venous Return* New York Grune & Stratton 1956

BRONK, D W AND G J STELLA Afferent impulses in carotid sinus nerve relation of discharge from single end organs to arterial blood pressure *J Cell Comp Physiol* 1 113-130 1932

BROWN A M Motor innervation of the coronary arteries of the cat *J Physiol* 198 311-328 1968

BRUNS D L A general theory of the causes of murmurs in the cardiovascular system *Amer J Med* 27 360-374 1959

BUGGE ASPERHEIM B, AND F KIIL Cardiac response to increased aortic pressure *Scand J Clin Lab Invest* 24 345-360 1969

BULLER, A J, W F H M MOMMAERTS, AND K SERAY DARIAN Enzymatic properties of myosin in fast and slow twitch muscles of the cat following cross innervation *J Physiol* 205 581-597 1969

BURCH, G E AND T WINSOR *A Primer of Electrocardiography* Philadelphia Lea & Febiger 1966

BURNS, W H H B SCHUMAKER, JR AND N J GRIFFITH The total mechanical cardiac substitute *Prog Cardiovasc Dis* 12 302-311 1969

BURTON A C Relation of structure to function of the tissues of the wall of blood vessels *Physiol Rev* 34 619-642 1954

BURTON A C *Physiology and Biophysics of Circulation* Chicago Year Book 1965

BUTTERWORTH J S, AND C A POINDEXTER Short PR interval associated with a prolonged QRS complex *Arch Intern Med* 69 437-445 1942

CLAUSEN J P The effects of physical conditioning A hypothesis concerning circulatory adjustments to exercise *Scand J Clin Lab Invest* 24 305-313 1969

COFFMAN, J D AND D E GREGG Reactive hyperemia characteristics of the myocardium *Amer J Physiol* 199 1143-1149, 1960

COMROE, J H *Physiology of Respiration* Chicago Year Book, 1965

COURNAND, A Air and blood In *Circulation of the Blood Medicine and Ideas*, edited by A P Fishman and D W Richards England Oxford Univ Press 1964, pp 3-70

COURNAND A, AND H A RANGES Catheterization of the right auricle in man *Proc Soc Exper Biol Med* 46 462-466 1941

CRAIB W H A study of the electrical field surrounding active heart muscle *Heart* 14 71-109 1927

CRANEFIELD, P The force of contraction of extrasystoles and the potentiation of force of the postextrasystolic contraction a historical review *Bull N Y Acad Med* 41 419-427 1965

CULBERTSON, J W R W WILKINS F J INGELFINGER AND S E BRADLEY The effect of the upright posture upon hepatic blood flow in normotensive and hypertensive subjects *J Clin Invest* 30 305-311 1951

CUSHING H Some experimental and clinical observations concerning states of increased intracranial tension *Amer J Med Sci* 124 375-400, 1902

DALY I DEB, AND C HEBB *Pulmonary and Bronchial Vascular Systems* Baltimore Williams & Wilkins 1966

DALY I DEB, AND E B VERNEY Cardiovascular reflexes *J Physiol (Paris)* 61 268-274, 1926

DAWES, G S *Foetal and Neonatal Physiology A Comparative Study of the Changes at Birth* Chicago Year Book 1968

DEAL C P JR, AND H D GREEN Comparison of changes in mesenteric resistance following splanchnic nerve stimulation with responses to epinephrine and norepinephrine *Circ Res* 4 38-44 1956

DEAN, A L The movements of mitral cusps in relation to the cardiac cycle *Amer J Physiol* 40 206-217 1916

DEJOURS, P, W GAREY, AND H RAHN Comparison of ventilatory and circulatory flow between animals in various physiological conditions *Resp Physiol* 9 108-117, 1970

DETWEILER, D K., J W BUCHANAN, G F FREGIN AND J D HILL Heart In *The Beagle as an Experimental Dog*, edited by A C Anderson Ames Iowa State Univ Press, 1970, p 232-246

DITMER, D S, AND R M GREBE (eds) *Handbook of Circulation* Philadelphia Saunders, 1963

DODGE, A T, H SANDLER, W A BAXLEY, AND R R HAWLEY Usefulness and limitations of radiographic methods for determining left ventricular volume, *Amer J Cardiol* 18 10-24, 1966.

DOUTHEIL U AND K KRAMER Über die differenzierung kreislauf regulierender Reflexe aus dem linken Herz *Pfluger Arch* 269 114-129 1959

DUBOIS A B, AND R MARSHALL Measurements of pulmonary capillary blood flow and gas exchange throughout the respiratory cycle in man *J Clin Invest* 36 1566-1571 1957

DUDEL J AND W TRAUTWEIN Das Aktions Potential under Mechanogramm des Herzmuskels unter den Einfluss der Delinung *Cardiologia* 25 344-362 1954

DÜRRER, D R T VANDAM G E FREUD M S JANSE F L MEIJLER AND R C ARZBAECHER Total excitation of the isolated human heart *Circulation* 41 899-912 1970

EDMAN, K A The active state and the force velocity relationship in cardiac muscle In *Heart Failure* edited by H Reindell J Keul and E Doll Stuttgart Thieme 1968 p 133-138

EINTHOVEN W G FAHR AND A DEWAART On the direction and manifest size of the variations of potential in the human heart and on the influence of the position of the heart on the form of the electrocardiogram Translated by H E Hoff and P Sekelj *Amer Heart J* 40 163-211 1950

ENGELBERG, J Physiological regulation the steady state *Physiologist* 9 69-88, 1966

FEIGL E O Carotid sinus reflex control of coronary blood flow *Circ Res* 23 223-237 1968

FEIGL E O Parasympathetic control of coronary blood flow in dogs *Circ Res* 25 509-579 1969

FERRARIO C M J W McCUBBIN AND I H PAGE Hemodynamic characteristics of chronic experimental neurogenic hypertension in unanesthetized dogs *Circ Res* 24 911-922, 1969

FICK A *Mechanische Arbeit u Warmeentwicklung bei der Muskeltatigkeit* Leipzig F A Brockhaus, 1882

FINNERTY F A JR, L WITKIN AND J F FAZEKAS Cerebral hemodynamics during cerebral ischemia induced by acute hypotension *J Clin Invest* 33 1227-1232 1954

FLOREY E *An Introduction to General and Comparative Animal Physiology* Philadelphia Saunders, 1966

FLORKIN, N, AND H F BLUM Sur la teneur en proteines du sang et du liquide coelonique des invertebres *Arch Int Physiol* 38 353-364, 1934

FOLKOW B Range of control of the cardiovascular system by the central nervous system *Physiol Rev* 40 (Suppl 4) 93-100 1960

FOLKOW, B, O LUNDGREN, AND I WALLENTINE Studies on the relationship between flow resistance capillary filtration coefficient and regional blood volume in the intestine of the cat *Acta Physiol Scand* 57 270-283, 1963

FORSSMAN W I Die Sondierung des rechten Herzen *Klin Wschr* 8 2085-2087, 1929

FRANK, O Zur Dynamik des Herzmuskels *Z Biol* 32 370-437 1895

FRANKLIN, D L, R M ELLIS, AND R F RUSHMER Aortic blood flow in dogs during treadmill exercise *J Appl Physiol* 14 809-812 1959

FRANKLIN, D L R L VAN CITTERS, AND R F RUSHMER Balance between right and left ventricular output *Circ Res* 10 17-26 1962

FRY D L F W NOBLE, AND A J MALLOS An electric device for instantaneous and continuous computation of aortic blood velocity *Circ Res* 5 75-78, 1957

GAGER, L T *Hypertension* Baltimore Williams & Wilkins, 1930

GARREY, W E The nature of fibrillary contraction of the heart its relation to tissue mass and form *Amer J Physiol* 33 397-414 1914

GARREY W E Auricular fibrillation *Physiol Rev* 4 215-250, 1924

GAUER, O H Zentrales Blutvolumen, Herzleistung und flussigkeitshanshalt In *Herzinsuffizienz*, edited by H Reindell, J Keul, and E Doll Stuttgart Thieme 1968, 523-529

GAUER, O H, J P HENRY, AND C BEHN The regulation of extracellular fluid volume *Ann Rev Physiol* 32 547-595 1970

GAUER, O H, J P HENRY AND H O SIEKER Changes in central venous pressure after moderate hemorrhage and transfusion in man *Circ Res* 4 79-84, 1956

GEDDES, L A *The Direct and Indirect Measurement of Blood Pressure* Chicago Year Book, 1970

GOLDBERGER, E Unipolar Lead Electrocardiography Philadelphia Lea & Febiger, 1947

GOODYER A V N The effects of hypothermia and perfusion pressure on left ventricular function in the intact mammal *Amer J Cardiol* 10 205-212 1963

GOORMAGHTIGH N La presence de cellules musculaires dans la paroi des arterioles du rein et leur comportement dans l'hypertension renale *C R Soc Biol (Paris)* 132 465-467 1938

GRANGER H S AND A C GUYTON Autoregulation of the total systemic circulation following destruction of the central nervous system in the dog *Circ Res* 25 379-388 1969

GRANT R P Notes on the muscular architecture of the left ventricle *Circulation* 32 301-308 1965

GRANT R T Observations on local arterial reactions in rabbit's ear *Heart* 15 257-280 1930

GREEN G E S H STERTZER R B GORDON AND D A TICE Anastomosis of the internal mammary artery to the distal left anterior descending coronary artery *Circulation* 41 and 42 (Suppl 2) 79-84 1970

GREEN H D Circulation Physical principles In *Medical Physics* edited by O Glasser Chicago Year Book 1944 vol 2 pp 765-857

GREEN H D R N LEWIS N D NICKERSON AND A C HELLER Blood flow peripheral resistance and vascular tonus with observations on relationship between blood flow and cutaneous temperature *Amer J Physiol* 141 518-536 1944

GREEN H D S H LOCKSLEY J SEXTON AND C P DEAL Autonomic vasomotor responses in the canine hepatic arterial and venous beds *Amer J Physiol* 196 196-202 1959

GREEN H D K OTTIS AND T KITCHEN Autonomic stimulation and blockade on canine splenic inflow outflow and weight *Amer J Physiol* 198 424-428 1960

GREENFIELD A D M AND G C PATTERSON On the capacity and distensibility of the blood vessels of the human forearm *J Physiol (London)* 131 290-306 1956

GREENSPAN K AND R E EDMANDS The inotropic effects of digitalis In *Digitalis*, edited by C Fisch and B Surawicz New York Grune & Stratton 1969

GREGERSEN M I AND S CHIEN Blood volume In *Medical Physiology* edited by V B Mountcastle St Louis Mosby 1968 vol 1

GREGG D E *Coronary Circulation in Health and Disease* Philadelphia Lea & Febiger 1950

GREGG D E Some problems of the coronary circulation *Verh Deutsch Ges Kreislaufforsch* 21 22-38 1955

GREGG D E The coronary circulation *Encyclopedia of Cardiology* New York McGraw Hill Blakiston 1959 vol 1 chap 23

GREGG, D E Coronary collateral circulation symposium Introduction to the problem *Acta Cardiol (Brux)* Suppl 13 1-12 1969

GREGG D E, AND D C SABISTON Current research and problems of the coronary circulation *Circulation* 13 916-927 1956

GREGG D E D C SABISTON AND E O THEILEN Performance of the heart changes in left ventricular end diastolic pressure and stroke work during infusion and following exercise *Physiol Rev* 35 130-136 1955

GROVE D R ZAK K S NAIR AND V ASCHENBRENNER Biochemical correlates of cardiac hypertrophy IV Observations on the cellular organization of growth during myocardial hypertrophy in the rat *Circ Res* 25 473-485, 1969

GUTH L Trophic influences of nerve on muscle *Physiol Rev* 48 645-687 1968

GUTTMAN L, AND D WHITTERIDGE Effects of bladder distension on autonomic mechanisms after spinal cord injuries *Brain* 70 361-404 1947

GUYTON A C Determination of cardiac output by equating venous return curves with cardiac response curves *Physiol Rev* 35 123-129 1955

GUYTON A C A concept of negative interstitial pressure based on pressures in implanted perforated capsules *Circ Res* 12 399-414 1963

GUYTON A C Pulmonary alveolar capillary interface and interstitium Introduction in *The Pulmonary Circulation and Interstitial Space*, edited by A P Fishman and H H Hecht Chicago Univ of Chicago Press 1969

GUYTON A C D POLIZO AND G G ARMSTRONG Jr Mean circulatory filling pressure measured immediately after cessation of heart pumping *Amer J Physiol* 179 261-267 1954

HADDY, F J Vasomotion in systemic arteries, small vessels and [illegible] determined by direct pressure measurements *Ann Med* [illegible]

HADDY F J M REINHMAN AND J B SCOTT Jr Effect of change in temperature upon systemic small and large vessel resistance *Proc Soc* [illegible]

HADDY F J AND P P GILBERT The relation of arteriolar reflex to transmural pressure and resistance in small and large systemic vessels *Circ Res* 4 [illegible]

HAMILTON W F R A WOODBURY AND F VOGT Differential pressures in the vessels of the circulation of the anaesthetized dog *Amer J Physiol* 125 [illegible]

HAMLIN R I AND C R SMITH Effects of vagal stimulation on S-A and A-V nodes *Amer J Physiol* 213 560-568 1968

HAMLIN R L C R SMITH AND D I SMETZER Sinus arrhythmia in the dog *Amer J Physiol* 210 [illegible]

HAMMOND G L E R JOCA AND W G AUSTEN The nature of intercoronary arterial flow *Amer Heart J* 78 559-565 1969

HAMRICK L W Jr AND J D MYERS The effect of hemorrhage on hepatic blood flow and splanchnic oxygen consumption of the dog *Circ Res* 3 67-72 1955

HANSON J AND J LOWY Molecular basis of contractility in muscle *Br Med Bull* 21 264-271 1965

HECHT H H Electrical properties of cardiac fibers In *Cardiovascular Disorders* edited by A N Brest and S J Moyer Philadelphia F A Davis 1968 p 79-89

HEDLUND S G NYLIN AND O REGNSTROM The behaviour of the cerebral circulation during muscular exercise *Acta Physiol Scand* 54 316-324 1962

HEINEMANN H O AND A P FISHMAN Non respiratory function of the mammalian lung *Physiol Rev* 49 1-47 1969

HERD J A Overall regulation of the circulation *Ann Rev Physiol* 32 289-312 1970

HEYMANS C J J BOUCKAERT AND P REGNIERS Le sinus carotidien Paris Doin 1933

HEYMANS C AND E NEIL Reflexogenic Areas of the Cardiovascular System Boston Little Brown 1958

HILL A V Heat of shortening and dynamic constants of muscle *Proc Roy Soc (London) [Biol] Ser B* 126 136-195 1939

HINSHAW L B Arterial and venous pressure resistance relationships in perfused leg and intestine *Amer J Physiol* 203 271-274 1962

HIRCHE H J Discussion of catecholamine effects on coronary circulation In *Coronary Circulation and Energetics of the Myocardium* edited by G Marchette and B Taccardi Basel Karger 1967

HODGKIN A L AND A F HUXLEY A quantitative description of membrane current and its application to conduction and excitation in nerve *J Physiol (London)* 117 500-544 1952

HOFFMAN B F Effects of digitalis on electrical activity of cardiac fibers In *Digitalis Mechanisms of the Inotropic Effect of Digitalis* edited by C Fisch and B Surawicz New York Grune & Stratton 1969 p 93-109

HOFFMAN B F AND P F CRANEFIELD The physiological basis of cardiac arrhythmias *Amer J Med* 37 670-684 1964

HOLLAND W C AND S SABATINI SMITH In *Digitalis Mechanisms of the Inotropic Effect of Digitalis* edited by B Surawicz and C Fisch New York Grune & Stratton 1969 p 43-54

HOLSINGER J W A G WALLACE AND W C SEALY The identification and surgical significance of the atrial internodal conduction tracts *Amer Surg* 167 447-453 1968

HOLT J P E A RHODE AND H KINES Ventricular volumes and body weight in mammals *Amer J Physiol* 215 704-715 1968

HORT, W Strukturelle Analyse akat insuffizientes Hertzen In *Heart Failure*, edited by H Reindell J Keul, and E Doll Stuttgart Thieme 1968, p 6-11

HOWARD A R W F HAMILTON AND P DOW Limitations of the continuous infusion method for measuring cardiac output by dye dilution *Amer J Physiol* 175 173-177 1953

HOWELL W F AND F DONALDSON Jr Experiments upon the heart of the dog *Phil Trans Roy Soc B (London)* 35 271-274 1883

HUGHES, J M B J B GRACIER, J E MALONEY, AND J B WEST Effect of lung volume on the distribution of pulmonary blood flow in man *Resp Physiol* 4 58 1968

JACOB R Druck Volumen Zeitbeziehungen im Tierexperiment In

Herzinsuffizienz edited by H Reindell J Keul and E Doll Stuttgart Thieme p 458-467 1968

JAMES T N Anatomy of the conduction system of the heart In *The Heart* edited by J W Hurst and R B Logue Ed 2 New York McGraw Hill 1970a

JAMES T N Cardiac conduction system *Amer J Cardiol* 25 213-226 1970b

JAMES T N E S BEAN K F LAND AND E W GREEN Evidence for alpha receptor depressant activity in the heart *Amer J Physiol* 215 1366-1375 1968

JAMES T N AND SHERF L Ultrastructure of myocardial cells *Amer J Cardiol* 22 289-416 1968

JEWELL B R AND D R WILKIE An analysis of the mechanical components in frog s striated muscle *J Physiol (London)* 143 515-540 1958

JEWELL B R AND D R WILKIE The mechanical properties of relaxing muscle *J Physiol (London)* 152 30-47 1960

JOHNSON P C Myogenic nature of increase in intestinal vascular resistance with venous pressure elevation *Circ Res* 7 992-999 1959

JOHNSON P C Autoregulation of intestinal blood flow' *Amer J Physiol* 199 311-318 1960

JOHNSON P C Review of previous studies and current theories of autoregulation *Circ Res* 15 (Suppl 2) 2-9 1964

JOSE A D AND R R TAYLOR Autonomic blockade by propranolol and atropine to study intrinsic myocardial function in man *J Clin Invest* 48 2019-2031 1969

KATTUS A A AND D E GREGG Some determinants of coronary collateral blood flow in the open chest dog *Circ Res* 7 628-642 1959

KATZ A M Patterns of energy production and energy utilization in cardiac and skeletal muscle In *Factors Influencing Myocardial Contractility* edited by R D Tanz F Kavaler and J Roberts New York Academic Press 1967 p 401-415

KATZ A M Contractile proteins of the heart *Physiol Rev* 50 63-158 1970

KEITH A AND M W FLACK The form and nature of the muscular connections between the primary divisions of the vertebrate heart *J Anat Physiol* 41 172-189 1907

KELLER A D Ablation and stimulation of the hypothalamus Circulatory effects *Physiol Rev* 40 (Suppl 4) 116-134 1960

KELLERMEYER R W AND R C GRAHAM JR Kinins Possible physiologic and pathologic roles in man *New Eng J Med* 279 754-759 802-807, 859-866 1968

KELLY D T The hemodynamic effects of paired pacing of the myocardium in reversible acute heart failure in the canine *Amer Heart J* 77 81-88 1969

KETY S S Physiology of the cerebral circulation in man In *Circulation* (Proceedings of the Harvey Tercentenary Congress) edited by J McMichael Oxford Blackwell, 1958 p 324-340

KETY S S AND C F SCHMIDT The nitrous oxide method for the quantitative determination of cerebral blood flow in man Theory procedure and normal values *J Clin Invest* 27 476-483 1948

KETY S S H A SHENKIN, AND C F SCHMIDT The effects of increased intracranial pressure on cerebral circulatory functions in man *J Clin Invest* 27 493-499 1948

KING B D, L SOKOLOFF AND R L WECHSLER The effects of l epinephrine and l-nor epinephrine upon cerebral circulation and metabolism in man *J Clin Invest* 31 273-279, 1952

KIRKENDALL, W M A S BURTON F H EPSTEIN AND E D FREIS Recommendations on human blood pressure determination sphygmomanometers *Subcommittee Report, American Heart Association, Circulation* 36 980-988 1967

KLEIN O Zur Bestimmung des zirkulatorischen minutenvolumens beim Menschen nach dem Fickschen Prinzip *Munchen Med Wschr* 77 1311-1312 1930

KLEINERMAN J AND S M SANCETTA Effect of mild steady state of exercise on cerebral and general hemodynamics of normal untrained subjects (Proceedings of the 47th Annual Meeting of the American Society for Clinical Investigation, 1955) *J Clin Invest* 34 (abs) 945 1955

KLEINERMAN, J, S M SANCETTA AND D B HACKEL Effects of high spinal anesthesia on cerebral circulation and metabolism in man *J Clin Invest* 37 285-293, 1958

KLOCKE, F, G KAISER, J ROSS, AND E BRAUNWALD Studies on beta adrenergic receptor mechanisms in the coronary vascular bed (abstr) *Fed Proc* 23 125 1964

KOCH E Die Irradiation der pressoreceptorischen Kreislaufreflexe *Klin Wschr* 11 225-227 1932

KOCH WESER J AND J R BLINKS The influence of the interval between beats on myocardial contractility *Pharmacol Rev* 15 601-652 1963

KOLFF W J S D MOULOPOULOS C S KWAN GETT AND A KRAHOS Mechanical assistance to the circulation the principle and the methods *Prog Cardiovasc Dis* 12 243-270 1969

KOUWENHOVEN W B J R JUDE AND G G KNICKERBOCKER Closed chest cardiac massage J A M A *173* 1064-1067 1960

KRAMER J G AND T W TODD The distribution of nerves to the arteries of the arm with a discussion of the clinical value of results *Anat Rec* 8 243-255 1914

KRAMER K AND P DEETJEN Sodium reabsorbtion and oxygen consumption in the mammalian kidney *Proceedings of the Ist International Congress of Nephrology Geneve/Evian* 687-691 1961

KRAMER K K THURAU AND P DEETJEN Hamodynamik des Nierenmarks I Capillare Passagezeit Blutvolumen Durchblutung Gewebshamatokrit und O, Verbrauch des Nierenmarks in situ *Pflueger Arch* 270 251-269 1960

LAKS M M F MORADY AND H J C SWAN Canine right and left ventricular cell and sarcomere lengths after banding the pulmonary artery *Circ Res* 24 705-710 1969

LANDIS E M The capillary circulation In *Circulation of Blood Men and Ideas* edited by A P Fishman and D W Richards New York Oxford University Press 1964

LANDOWNE M Characteristics of impact and pulse wave propagation in brachial and radial arteries *J Appl Physiol* 12 91-97 1958

LEGATO M J The correlation of ultrastructure and function in the mammalian myocardial cell *Prog Cardiovasc Dis* 11 391-409 1969

LEGATO, M J Sarcomergenesis in the human myocardium *J Molec Cell Cardiol* 1 425-437 1970

LEVY M N, S H BRIND, F R BRANDLIN, AND F A PHILLIPS The relationship between pressure and flow in the systemic circulation of the dog *Circ Res* 2 372-380 1954

LEVY M N N G MIL AND H ZIESKE Functional distribution of the peripheral cardiac sympathetic pathways *Circ Res* 19 650-661 1966

LEWIS D W Phonocardiography In *Handbook of Physiology* edited by W F Hamilton and P Dow Washington Amer Physiol Soc 1962 sect 2 vol 1, p 695-734

LEWIS T *The Mechanism and Graphic Registration of the Heart Beat* Ed 3 London Shaw and Sons 1925

LEWIS T *The Blood Vessels of the Human Skin and Their Responses* Shaw, London 1927

LEWIS T, AND R GRANT Observations upon reactive hyperaemia in man *Heart* 12 73-120 1925

LIEBERMAN, M Physiologic development of impulse conduction in embryonic cardiac tissue *Amer J Cardiol* 25 278-284 1970

LINDGREN P, AND B UVNAS Vasoconstrictor inhibition and vasodilator action—two functionally separate vasodilator mechanisms in the skeletal muscles *Acta Physiol Scand* 33 108-119 1955

LING G AND R W GERARD The normal membrane potential of frog sartorious fibers *J Cell Comp Physiol* 34 383-396 1949

LINZBACH A J Heart failure from the point of view of quantitative anatomy *Amer J Cardiol* 5 370-382, 1960

LINZBACH J AND C KYRIELEIS Strukturelle Analyse chronisch insuffizienter menschlicher Herzen In *Heart Failure*, edited by H Reindell H Keul and E Doll Stuttgart Thieme 1968 p 11-19

LIPMAN, B S AND E MASSIE *Clinical Scalar Electrocardiography* Ed 4 Chicago Year Book 1959

LITTER, J, AND J E WOOD, JR The venous pressure volume curve of the human leg measured in vivo (abstr) *J Clin Invest* 33 953, 1954

LOPEZ MUNIZ R, N L STEPHENS B BROMBERGER BAREA S PERMUTT, AND R L RILEY Critical closure of pulmonary vessels analyzed in terms of Starling resistor model *J Appl Physiol* 24 625-635 1968

LOWN B Cardioversion of arrhythmias *Modern Conc Cardiovasc Dis*, 33 863-868 1964

MACKENZIE, J *Diseases of the Heart* London Frowde 1913

MALONEY J E D H BERGEL J B GLAZIER J M B HUGHES AND J B WEST Transmission of pulsatile blood pressure and flow through the isolated lung *Circ Res* 21 11-24 1967

MANASEK F J Histogenesis of the embryonic myocardium *Amer J Cardiol* 25 149-168 1970

MANGOLD R I SOKOLOFF E CONNER J KLEINERMAN P G THERMAN AND S S KETY The effects of sleep and lack of sleep on the cerebral circulation and metabolism of normal young men *J Clin Invest* 34 1092-1100 1955

MARRIOTT H J L Nodal mechanisms with dependent activation of atria and ventricles In *Mechanisms and Therapy of Cardiac Arrhythmias* edited by L S Dreifus and W Likoff New York Grune & Stratton 1966 pp 412-418

MARRIOTT H J L AND R J MYERBURG Cardiac arrhythmias and conduction disturbances In *The Heart* edited by J W Hurst and R B Logue New York McGraw Hill 1970 p 489-543

MAYER A G Papers from Tortugas Lab Carnegie Inst Washington 1 115 1908

MAYER S E Adrenergic receptors for metabolic responses in the heart *Fed Proc* 29 1367-1372 1970

MAYERSON H S Effect of gravity on the blood pressure of the dog *Amer J Physiol* 135 411-418 1942

McCUBBIN J W C J DICKINSON W M MANGER AND P KEZDI Neurogenic factors In *Renal Hypertension* edited by I H Page and J W McCubbin Chicago Year Book 1968 p 244-262

McNUTT N S Ultrastructure of intercellular junctions in adult and developing cardiac muscle *Amer J Cardiol* 25 169-183 1970

McWILLIAM J A Fibrillar contraction of the heart *J Physiol* 8 296-310 1887

MEERSON F Z The myocardium in hyperfunction hypertrophy and heart failure *Circ Res* 25(Suppl 2) 1-163 1969

MELLANDER S Systemic circulation local control *Ann Rev Physiol* 32 313-344 1970

MELLANDER S AND B JOHANSSON Control of resistance exchanges and capacitance functions in the peripheral circulation *Pharmacol Rev* 20 117-196 1968

MILHORN H T JR *Application of Control Theory to Physiological Systems* Philadelphia Saunders 1966

MILHORN H T JR Practical Application of control theory in respiratory and circulatory research *Fed Proc* 28 84-88 1969

MILLER F R AND J T BOWMAN The cardioinhibitory center *Amer J Physiol* 39 149-153 1915

MILNOR W R Pulmonary circulation In *Medical Physiology* ed ited by V B Mountcastle St Louis Mosby 1968 vol 1 p 24-34 and 209-220

MINES G R On dynamic equilibrium in the heart *J Physiol (London)* 46 349-383 1913

MITCHELL J AND W SHAPIRO Atrial function and the hemodynamic consequences of atrial fibrillation in man *Amer J Cardiol* 23 556-557 1969

MOE G K W C RHEINBOLDT AND J A ABILDSKOV A computer model of atrial fibrillation *Amer Heart J* 67 200-220 1964

MOORE, E N, S KNOEBEL AND J SPEAR Concealed conduction *Amer J Cardiol* 28 406-413 1971

MORKIN E J A COLLINS H S GOLDMAN AND A P FISHMAN Pattern of blood flow in the pulmonary veins of the dog *J Appl Physiol* 10 1118-1128 1965

MOYER, J H, S I MILLER AND H SNYDER Effect of increased jugular pressure on cerebral hemodynamics *J Appl Physiol* 7 245-247 1954

MURPHY Q The influence of the accelerator nerves on the basal heart rate of the dog *Amer J Physiol* 137 727-730 1942

NAHAS, G G, E C JORDAN, AND J C LIGOU Effects of a CO_2 buffer' on hypercapnia of apneic oxygenation *Amer J Physiol* 197 1308-1316 1959

NELSON D, AND J F FAZEKAS Cerebral blood flow in polycythemia vera *Arch Intern Med* 98 328-331, 1956

NOER R J H J ROBB AND L F JACOBSON Circulatory disturbances produced by acute intestinal distension in the living animal *Arch Surg* 63 520-528 1951

NONIDEZ, J F The structure and innervation of the conductive system of the heart of the dog and rhesus monkey as seen with a silver impregnation technique *Amer Heart J* 26 577-597 1943

OLSON R E AND D A PIATNEK Conservation of energy in cardiac muscle *Ann N Y Acad Sci* 72 466-478 1959

OLSON R E N S DHALLA AND C N SUN Changes in energy stores in the myocardium *Circ Res* 56 114-124 1971 72

ORMSBY O S AND H MONTGOMERY *Diseases of the Skin* Philadelphia Lea & Febiger 1954

PARSON W H S MAYERSON C LYONS B PORTER AND W V TRAUTMAN The Effect of the administration of adrenalin on the circulation of the dog *Amer J Physiol* 155 239-241 1948

PASCARELLI F E AND C A BERTRAND Comparison of blood pressures in the arms and legs *N E J Med* 270 691-695 1964

PATTERSON D F D K DETWEILER AND S A GLENDINNING Heart sounds and murmurs of the normal horse *Ann N Y Acad Sci* 127 214-265 1965

PATTERSON J L JR Cardiorespiratory dynamics in the ox and giraffe with comparative observations on man and other animals *Ann N Y Acad Sci* 127 29-41 1965

PATTERSON J L A HEYMAN I I BATTEY AND R W FERGUSON Threshold response of the cerebral vessels of man to increase in blood carbon dioxide *J Clin Invest* 34 1857-1864 1955

PATTERSON S W H PIPER AND E H STARLING The regulation of the heartbeat *J Physiol (London)* 48 465-513 1914

PERMUTT S AND R L RILEY Hemodynamics of collapsible vessels with tone the vascular waterfall *J Appl Physiol* 18 924-932 1963

POCHE R Ultrastructure of heart muscle under pathological conditions *Ann N Y Acad Sci* 156 34-47 1969

POLLACK G H Maximum velocity as an index of contractility in cardiac muscle A critical evaluation *Circ Res* 26 111 127 1970

POLLACK G H Is V max a valid contractile index? *Amer Heart J* 81 572-575 1971

POOL P E AND F H SONNENBLICK The mechanochemistry of cardiac muscle *J Gen Physiol* 50 931-965 1967

PRINZMETAL M F CORDAY I C BRILL R W OBLATH AND H E KRUGER The auricular arrhythmias Springfield Illinois Thomas 1952

PRIOLA D V Individual chamber sensitivity to norepinephrine after unilateral cardiac denervation *J Physiol (London)* 216 604-614 1969

PROCTOR E Assisted circulation in the experimental preparations —with special reference to pulmonary embolism and acute myocardial ischemia *Prog Cardiovasc Dis* 12 271-292 1969

RABINOWITZ M AND R ZAK Biochemical and cellular changes in cardiac hypertrophy *Ann Rev Med* In press

RAPAPORT E Usefulness and limitation of thermal washout techniques in ventricular volume measurement *Amer J Cardiol* 18 226-230 1966

RASMUSSEN H Cell communication calcium ion and cyclic adenosine monophosphate *Science* 170 404-412 1970

REEVE E B AND A C GUYTON (Eds) *Physical Basis of Circulatory Transport Regulation and Exchange* Philadelphia Saunders 1967

REMINGTON J W C R NORBACK W F HAMILTON AND J J GOLD Volume elasticity characteristics of the human aorta and prediction of the stroke volume from the pressure pulse *Amer J Physiol* 153 298 308 1948

REYNELL P C T A MARKS C CHIDSEY AND S E BRADLEY Changes in splanchnic blood volume and splanchnic blood flow in dogs after hemorrhage *Clin Sci* 14 407-419 1955

RIGGS D S *The Mathematical Approach to Physiological Problems* Baltimore Williams & Wilkins 1963

RIGGS D S *Control Theory and Physiological Feedback Mechanism* Baltimore Williams & Wilkins 1970

ROBB J S AND R S ROBB The normal heart *Amer Heart J* 23 455-467 1942

ROBIN E D, AND F H GARDNER Cerebral metabolism and hemodynamics in pernicious anemia (Proceedings of the 45th Annual Meeting of the American Society for Clinical Investigation 1953 *J Clin Invest* 32 (abs) 598 1953

ROSS J, JR, C J FRAHM AND E BRAUNWALD Influence of intracardiac baroreceptors on venous return, systemic vascular volume and peripheral resistance *J Clin Invest* 40 563-572 1961a

ROSS J JR C J FRAHM AND E BRAUNWALD The influence of the carotid baroreceptors and of vasoactive drugs on systemic vascular volume and venous distensibility *Circ Res* 9 75-82 1961b

ROWE G G, G M MAXWELL C A CASTILLO, D J FREEMAN, AND C W CRUMPTON A study in man of cerebral blood flow and cerebral glucose lactate and pyruvate metabolism before and after eating *J Clin Invest* 38 2154-2158 1959

RUSHMER, R F *Cardiovascular Dynamics* Ed 2 Philadelphia Saunders 1961

RUSHMER R F *Cardiovascular Dynamics* Ed 3 Philadelphia Saunders 1970

SABISTON D C AND D E GREGG Effect of cardiac contraction on coronary blood flow *Circulation* 15 14-20 1957

SARNOFF S J Certain aspects of the role of catecholamines in circulatory regulation *Amer J Cardiol* 5 579-588 1960

SARNOFF S J AND E BERGLUND Pressure volume characteristics and stress Relaxation in the pulmonary vascular bed of the dog *Amer J Physiol* 171 238-244 1952

SARNOFF S J AND E BERGLAND The Potter electroturbinometer *Circ Res* 1 331-336 1953

SARNOFF S J R B CASE G H WELCH JR E BRAUNWALD AND W N STAINSBY Performance characteristics and oxygen debt in a nonfailing metabolically supported isolated heart preparation *Amer J Physiol* 192 141-147 1958

SCHEINBERG P AND E A STEAD, JR The cerebral blood flow in male subjects as measured by the nitrous oxide technique Normal values for blood flow oxygen utilization glucose utilization and peripheral resistance with observations on the effect of tilting and anxiety *J Clin Inv* 28 1163-1171 1949

SCHERF D Versuche zur Theorie des Vorhofflatterns und Vorhofflimmerns *Z Ges Exp Med* 61 30-44 1928

SCHEUER J Metabolism of the heart in cardiac failure *Prog Cardiovasc Dis* 13 24-54 1970

SCHIEVE J F AND W P WILSON The changes in cerebral vascular resistance of man in experimental alkalosis and acidosis *J Clin Invest*, 32 33-38 1953

SCHLANT R C Metabolism of the heart In *The Heart*, edited by J W Hurst and R B Logue Ed 2 New York McGraw Hill 1970 p 111-113

SCHULTZ D L D S TUNSTALL PEDOE G DE J LEE A J GUNNING AND B J BELLHOUSE Velocity distribution and transition in the arterial system In *Circulatory and Respiratory Mass Transport, a Ciba Foundation Symposium*, edited by G E W Wolstenholme and Julie Knight Boston Little Brown 1969

SCHÜTZ, E *Lehrbuch der Physiologie, Physiologie des Herzens* Berlin Springer Verlag 1958

SCHWEITZER A Rhythmical fluctuations of the arterial blood pressure *J Physiol (London)* 104 (abs) 25 1945

SELKURT E E Intestinal ischemic block and the protective role of the liver *Amer J Physiol* 197 281-285 1959

SELKURT E E AND G A BRECHER, Splanchnic hemodynamics and oxygen utilization during hemorrhagic shock in the dog *Circ Res* 4 693-704 1956

SELKURT, E E AND R W BULLARD The heart as a pump In *Physiology*, edited by E E Selkurt Boston Little Brown 1966

SELKURT E E AND P C JOHNSON Effect of acute elevation of portal venous pressure on mesenteric blood volume interstitial fluid volume and hemodynamics *Circ Res* 6 592-599 1958

SHENKIN, H A SHEUERMAN, W G SPITZ, E B, AND GROFF R A The effects of change of position upon the cerebral circulation of man (Proceedings of the Physiological Society of Philadelphia 1948) *Amer J Med Sci* 216 (abs) 714-715 1948

SHIPLEY R E D E GREGG, AND E F SCHROEDER An experimental study of flow patterns in various peripheral arteries *Amer J Physiol* 138 718-730 1943

SHIPLEY R E AND C WILSON An improved recording rotameter *Proc Soc Exp Biol Med* 78 724-728 1951

SIDKY M, AND J W BEAN Influence of rhythmic and tonic contraction of intestinal muscle on blood flow and blood reservoir capacity in dog intestine *Amer J Physiol* 193 386-392, 1958

SMYTHE, C M, J F NICKEL, AND S E BRADLEY The effect of l-norepinephrine on estimated hepatic blood flow in the normal anesthetized dog (Proceedings of 45th Animal Meeting of the American Society for Clinical Investigation 1953) *J Clin Invest* 32 (abs.) 605, 1953

SNOOK, T A comparative study of the vascular arrangements in mammalian spleens. *Amer J Anat.* 87: 31-61, 1950

SOBIN S S H M TREMER AND Y C FUNG Morphometric basis of the sheet flow concept of the pulmonary alveolar microcirculation *Circ Res* 26 397-414 1970

SOKOLOFF L R MANGOLD R L WECHSLER C KENNEDY AND S S KETY The effect of mental arithmetic on cerebral circulation and metabolism *J Clin Invest*, 34 1101-1108 1955

SOMMER J R AND E A JOHNSON Comparative ultrastructure of cardiac cell membrane specialization A Review *Amer J Cardiol* 25 278-284 1970

SONNENBLICK E H The mechanics of myocardial contraction In *The Myocardial Cell* edited by S A Briller and H L Conn Jr Philadelphia Univ of Pennsylvania Press 1966 p 173-250

SONNENBLICK E H Correlation of myocardial ultrastructure and function *Circulation* 38 29-44 1968

SONNENBLICK, E H Contractility of cardiac muscle (letter to editor) *Circ Res* 27 479-481, 1970

SONNENBLOCK E H W W PARMLEY C W URSCHEL AND D L BRUTSAERT Ventricular Function Evaluation of myocardial contractility in health and disease *Prog Cardiovasc Dis* 12 449-466 1970

SOROFF A S F GERON A RINZ W C BERTWELL L J HIRSCH AND R A DETERLING Physiologic support of heart action *N Eng J Med* 280 693-704 1969

SPANN J F JR R A BUCCINO E H SONNENBLICK, AND E BRAUNWALD Contractile state of cardiac muscle obtained from cats with experimentally produced ventricular hypertrophy and heart failure *Circ Res* 21 341-354 1967

SPENCER, M P F R JOHNSTON AND A B DENISON Dynamics of the normal aorta *Circ Res* 6 491-500 1958

SPIRO D AND E H SONNENBLICK Comparison of contractile process in heart and skeletal muscle *Circ Res* 15(Suppl 2) 14-37 1964

SPOTNITZ H M E H SONNENBLICK AND D SPIRO Relation of ultrastructure to function in intact heart sarcomere structure relative to pressure volume curves of intact left ventricles of dog and cat *Circ Res* 18 49-66 1966

STARLING E H AND M B VISSCHER Regulations of energy output of the heart *J Physiol (London)* 62 243-261, 1926-1927

STEGEMANN, J AND A TIBES Sinusoidal stimulation of carotid sinus baroreceptors and peripheral blood pressure in dogs *Ann N Y Acad Sci* 156 787-795 1969

STRICKER, S (abstr III) *S B Akad Wiss Wien* 74 173 1876

TEXTER E C JR S MERRILL M SCHWARTZ G VAN DER STAGEN AND F J HADDY Relationship of blood flow to pressure in the intestinal vascular bed of the dog *Amer J Physiol* 202 253-256 1962

THORBURN G D, H H KOPALD, J A HERD M HOLLENBERG, C C C O'MORCHOE AND A C BARGER Intrarenal distribution of nutrient blood flow determined with krypton 85 in the unanesthetized dog *Circ Res* 13 290-307 1963

THURAU, K Renal hemodynamics *Amer J Med* 36 698-719 1964

THURAU, K, AND K KRAMER Weitere Untersuchungen zur myogenen Natur der Autoregulation des Nierenkreislaufes Aufhebung der Autoregulation durch muskulotrope Substanzen und druckpassives Verhalten des Glomerulusfiltrates *Pflueger Arch* 269 77-93 1959

THURAU, K AND J SCHNERMANN Die Natriumkonzentration an den macula Densa Zellen als regulierender Faktor fur das Glomerulumfiltrat (Mikropunktionsversuche) *Klin Wschr* 43 410-413 1965

TICE, L *In situ* studies of the sarcotubular system In *Myocardial Contractility* edited by R D Tanz, F Kavaler, and J Roberts New York Acad Press, 1967 pp 339-350

TORRANCE H B The control of the hepatic arterial circulation *J Physiol* 158 39-49, 1961

TRAUTWEIN, W Generation and conduction of impulses in the heart as affected by drugs *Pharmacol Rev* 15 277-332 1963

TRUEX R C Anatomy of the heart In *Encyclopaedia Britannica* Chicago Encyclopaedia Britannica 1967 vol 11, p 220-222

TSAKIRIS A G R A VANDENBERG, N BANCHERO, R E STURM AND E H WOOD Variations of left ventricular end diastolic pressure volume and ejection fraction with changes in outflow resistance in anesthetized dogs *Circ Res* 23 213-22, 1968

TURK L N III AND W C SHOEMAKER Hepatic vascular response to norepinephrine *Amer J Physiol* 202 1173-1178 1962

TURNER J C J LAMBERTSEN S G OWEN H WENDEL AND H CHIODI Effects of 06 and 8 atmospheres of inspired pO upon cerebral hemodynamics at a constant alveolar pCo of 4 mm Hg *Fed Proc* 16 (abst) 140 1957

TYBERG J V, W J KEON E H SONNENBLICK, AND C W URSCHEL Mechanics of ventricular diastole *Cardiovasc Res* 4 423-428 1970

VAN CITTERS R L AND D L FRANKLIN Cardiovascular research in wild animals Telemetry technics and applications *Acta Zool Path Antverp* 48 243-263 1969a

VAN CITTERS R L AND D L FRANKLIN Cardiovascular responses in Alaska sled dogs during exercise *Circ Res* 24 33-42 1969b

VASSALLE M AND B F HOFFMAN Spread of sinus activation during potassium administration *Circ Res* 17 285 295 1965

VATNER, S F D FRANKLIN R L VAN CITTERS AND E BRAUNWALD Effects of carotid sinus nerve stimulation on blood flow distribution in conscious dogs at rest and during exercise *Circ Res* 27 495-503 1970

WADE O L B L COMBES A W CHILDS H O WHEELER S A COURNAND, AND S E BRADLEY The effect of exercise on the splanchnic blood flow and splanchnic blood volume in normal man *Clin Sci* 15 457-463 1956

WAGNER M L R LAZZARA R M WEISS AND B F HOFFMAN Specialized conducting fibers in the interatrial band *Circ Res* 18 502-518 1966

WASSERMAN A J AND J L PATTERSON JR The cerebral vascular response to reduction in arterial carbon dioxide tension *J Clin Invest* 40 1297-1303 1961

WEIDMANN, S Effect of current flow on the membrane potential of cardiac muscle *J Physiol (London)* 115 227-236, 1951

WEIDMANN, S Diffusion of radio potassium across intercalated discs of mammalian cardiac muscle *J Physiol (London)* 187 323-342 1966

WETTERER, E Flow and pressure in the arterial system Their hemodynamic relationship and the principles of their measurement *Minn Med* 37 77-86 1954

WHIPPLE A O A K PARPART AND J J CHANG A study of the circulation of the blood in the spleen of the living mouse *Ann Surg* 140 266-283 1954

WHITMORE R L *Rheology of the Circulation* Oxford Pergamon Press 1968

WIGGERS C J *Circulatory Dynamics* New York Grune & Stratton 1952

WIGGERS C J *Physiology in Health and Disease* Ed 5 Philadelphia Lea & Febiger 1954

WILKINS R W J W CULBERTSON AND F J INGELFINGER The effect of splanchnic sympathectomy in hypertensive patients upon estimated hepatic blood flow in the upright as contrasted with the horizontal position *J Clin Invest* 30 312-317 1951

WILSON F N *Selected Papers* edited by F D Johnson and E Lepeschkin Ann Arbor Heart Station University Hospital 1954

WIRZ H B HARGITAY AND W KUHN Lokalisation des Konzentrierungsprozesses in der Niere durch direkte Kryoskopie *Helv Physiol Pharmacol Acta* 9 196 207 1951

WOLF A V Demonstration concerning pressure tension relations in various organs *Science* 115 243 244 1952

WOOD F C C C WOLFERTH AND G D GECKELER Histologic demonstration of accessory muscular connections between auricle and ventricle in case of short P R interval and prolonged QRS complex *Amer Heart J* 25 454-462 1943

WOOD P *Diseases of the heart and circulation* Ed 2 London Eyre and Spleswode 1968

WOODBURY L A H H HECHT AND A R CHRISTOPHERSON Membrane resting and action potentials of single cardiac muscle fibers of the frog ventricle *Amer J Physiol* 164 307-318 1951

YAMAMOTO W S AND J R BROBECK (Eds) *Physiological Controls and Regulations* Philadelphia Saunders 1965

YOUMANS W B Q R MURPHY L N DAVIES V I BRIGGS AND A S HOYE *The Abdominal Compression Reaction* Baltimore Williams & Wilkins 1963

ZERAHN K Oxygen consumption and active sodium transport in the isolated and short circuited frog skin *Acta Physiol Scan* 36 300-318 1956

ZUBERBUHLER R, AND D F BOHR Effects of catecholamines and angiotensin on coronary vascular smooth muscle (abstr) *Fed Proc* 23 121 1964

4

Blood and Lymph

PREFACE

In this section on the blood and lymph these fluids will be treated in respect to their transport of nutrients, waste products, chemical communications system, cellular and humoral defense system, and the means by which the milieu intérieur of Claude Bernard is maintained. In developing the function of the blood and lymph which interconnects all tissues of the body through the vascular system, the regulation of the fluctuating steady-state systems through feedback loops will be considered. Many of these feedback loops are considered in detail in the other chapters involving the kidney, lungs, brain, and endocrines. Only those that relate directly to those elements of the blood will be considered. The general plan of attack will focus on the functions related to the red cell, the granulocytic leukocytes, the lymphoreticular system, megakaryocytic-platelet system, the proteins related to blood coagulation and hemostasis, and the humoral aspects of immunity. The nature of the steady-state cell proliferation will be developed conceptually with reference to how it is studied and controlled.

BLOOD AND LYMPH

Blood: Composition and Functions

Phylogenetic and ontogenetic study of species indicates that life commenced in some simple form in a primitive sea, and evolved from single organisms to the complex Metazoa Single cell organisms such as Protozoa and bacteria exist primarily in a fluid medium Such organisms acquire small molecules required for metabolism by direct diffusion from the environment through membranes and into the cell Diffusion is a two-way street with products of metabolism diffusing outward In some, if not all organisms, in addition to diffusion, material is ingested by phagocytosis and pinocytosis The ingested material is incorporated into phagocytic vacuoles into which hydrolytic and oxidative enzymes are injected breaking down offensive material or particulate nutrients to simple products that can then be either excreted or utilized in the general metabolism of the cell

As organisms progress from single cell to multicellular organisms with complex structure, diffusion, pinocytosis, and phagocytosis are inadequate to provide the small molecular needs of each cell Tubules, propulsion systems of cilia, and ultimately a pump with a closed circulatory system develops and transports nutrients and waste products As complex structures that break down incoming food develop, the biochemical, structural, and transport systems become increasingly complex Each segment of the increasingly complex system undertakes specific and multiple functions and so maintains the internal environment within the range permissible and compatible with life irrespective of the external environment

In the mammal, the blood performs various specific functions

1) Respiratory. An essential function is the transport of oxygen from air in the lungs to the tissues, and the carbon dioxide from the tissues to the lungs

2) Nutritive. Blood conveys food materials, glucose, amino and fatty acids, vitamins, electrolytes, and trace metals from the alimentary canal to the tissues

3) Excretory Blood transports the waste products of metabolism such as urea, uric acid, creatinine, etc

4) Homeostasis of Water, pH, and Electrolyte Concentration. There is constant exchange of all molecules in the body between blood, interstitial fluid and cells These exchanges are in dynamic equilibrium With the exception of deoxyribonucleic acid (DNA) in the nondividing cells, all constituents of cells are in a continual steady-state turnover at different rates This turnover is mediated through the blood being in direct contact with cells through the interstitial fluid and lymph

5) Regulation of Body Temperature A major function of the blood is heat transport This is essential to the preservation of the very narrow range in body temperatures that is compatible with life This is considered in greater detail elsewhere in this text Water is pre-eminently fitted to fulfill the purpose of heat transport

1) High Specific Heat The specific heat of a substance is defined as the number of calories required to raise 1 g of the substance 1° C The specific heat of water is much higher than that of any other liquid or solid, and this buffers sudden changes in temperature A man of average weight generates 3000 kcal within 24 hr This amount of heat, if it were all retained in the body, is capable of raising the temperature of his tissues, which are 80% water, only about 32° C. Heat elimination (radiation, etc) normally dissipates this heat production, and the body temperature varies only slightly within narrow limits If the tissues had a low specific heat of most substances, 3000 kcal would raise the temperature of the body by 100 to 150° C (Henderson, 1913)

2) High Conductivity The thermal conductivity of water is greater than that of any other ordinary liquid The advantage of this in heat transport from deep organs to the surface,

even without the circulation of blood, is obvious

3) High Latent Heat of Evaporation More heat is required for the vaporization of water than for any other liquid Fluid is being constantly evaporated from the lungs and skin Through this process a large amount of heat is lost

The circulation of blood transporting heat from within the body to the lungs, the surface and the various vascular shunts that increase and decrease the circulation through the skin serves to accelerate the loss of heat or to retain the heat when needed

6) Chemical Communication and Protection. In addition to the nervous system that communicates from the brain to the various organs, there are sensitive chemical homeostatic systems The concentration of various substances in the blood is regulated through feedback loops that sense changes in concentration and trigger the synthesis of hormones and enzymes which initiate the synthesis of substances needed in various organs Furthermore, the entire complex of humoral antibodies important in defense against infection, initiation of inflammation, and the regulation of hemostasis circulate within the blood These will be considered in detail in other chapters

The Composition of Blood

Blood is a complex non-newtonian fluid in which several different types of cells are suspended The specific gravity varies among individuals from 1 050 to 1 060 and its viscosity is roughly 5 to 6 times that of water If blood clotting is prevented by addition of substances that bind calcium or inhibit later stages in the coagulation process, the cellular elements can be spun down by centrifugation The supernatant material is called *plasma* and contains a wide spectrum of proteins, organic and inorganic substances in solution, hormones, antibodies, and substances not yet adequately defined The specific gravity of plasma is roughly 1 027 but varies with the concentration of protein in health and disease over wide ranges In health, the cells constitute about 46% by volume of human blood, and the plasma 54%

If blood is permitted to clot, the fibrin mesh traps the cells and retracts When this is centrifuged the supernatant material is called *serum* and is identical to plasma with the exception of the removal of fibrinogen and other clotting factors

Whole blood
- *A)* Cells
  - *1)* Red corpuscles (erythrocytes)
  - *2)* White corpuscles (leukocytes)
  - *3)* Platelets (thrombocytes)
- *B)* Plasma
  - *1)* Water, 91 to 92%
  - *2)* Solids, 7 to 9%
    - *a)* Proteins, 7% Serum albumin, globulins, and fibrinogen
    - *b)* Inorganic constituents, 0 9% Sodium, calcium, potassium, magnesium, phosphorus, iodine, iron, copper, etc
    - *c)* Organic constituents (other than *a)* and *d)* Nonprotein nitrogenous substances, (urea, uric acid, xanthine, hypoxanthine, creatine ~~*a)* and *d*)~~ Nonprotein nitroge amino acids) neutral fats, phospholipids, cholesterol, glucose
    - *d)* Internal secretions, antibodies, and various enzymes, (amylases, proteases, lipases, esterases, etc)

INORGANIC CONSTITUENTS

Plasma is relatively rich in sodium and calcium but poor in potassium and magnesium (Table 4 1), whereas the reverse is true in cells These gradients are maintained by dynamic energy-consuming processes All of the constituents are in a steady-state exchange that can be shown only by the movement of radioactively labeled materials The iron in cells is confined to hemoglobin and porphyrin enzymes The iron and copper of plasma are transported and bound by specific metal binding proteins These play a very important role in hemopoiesis

PHOSPHORUS

Phosphorus exists in blood in four main forms Most of it is present as organic esters in the red cells The importance of these is considered in chapter 4 The phosphorus fractions in organic combinations are

1) Ester phosphorus, e g , diphosphoglycerate, adenosinetriphosphate, hexose phosphates, glycerophosphate and a whole host of red cell enzymes,

Table 4-1
Chemical composition of peripheral lymph (cervical) and blood plasma from the dog

	Protein (Kjeldahl)	Non-protein nitrogen	Urea	Uric acid	Creatinine	Sugar	Amino acids	Chlorides as NaCl	Phosphorus		Calcium
									Total	Inorganic	
	%	mg / 100 ml	mg / 100 ml	mg / 100 ml	mg / 100 ml	mg / 100 ml	mg / 100 ml	mg / 100 ml	mg / 100 ml	mg / 100 ml	mg / 100 ml
Plasma											
Average	6 18	32 6	21 7	Tr	1 37	120 0	4 90	678	22 0	5 6	11 70
Range	5 54-	21 1-	17 9-		1 22-	112 0-		649-	18 3-	4 4-	10 85-
	7 23	46 0	28 0		1 54	143 0		721	26 1	6 9	12 95
Lymph											
Average	3 92	34 8	23 5	Tr	1 40	132 2	4 84	711	11 8	5 9	9 84
Range	1 38-	19 8-	19 8-		1 28-	107 0-		690-	10 2-	4 7-	8 93-
	4 57	45 4	33 0		1 49	144 0		730	13 7	7 3	10 84
No of animals	16	10	7	3	7	16	1	7	6	3	11

From Heim 1933
[a] Tr , trace

2) *Lipid phosphorus*, e g , the phosphatides lecithin, cephalin, sphingomyelin

3) *Nucleic acid phosphorus* is distributed between deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) The former is almost exclusively in nuclei as the genetic material of the chromosomes, although a tiny amount of DNA is in mitochondria of the cytoplasm RNA is distributed between nuclei and cytoplasm in the form of messenger RNA, transfer and ribosomal RNA, the chemical messengers, and the protein synthesizing machinery

4) *Inorganic phosphorus* is distributed between cells and plasma, and is considered in other chapters

The inorganic and ester fractions are extracted from blood by the precipitation of the proteins with trichloracetic acid and filtration The phosphorus contained in the filtrate is the acid soluble phosphorus Upon extraction of blood with alcohol-ether the lipid phosphorus is obtained The phosphorus of blood is therefore separated into two classes

1) The acid soluble fraction which includes
   a) Inorganic phosphorus
   b) Ester phosphorus
2) The alcohol-ether soluble, i e , lipid phosphorus

The ester and lipid phosphorus constitute the organic phosphorus of blood

The organic phosphorus compounds are exceptionally important in energy transfer processes considered in Sect 1, Chap 4

PLASMA PROTEINS

The total concentration of proteins in the plasma and the proportions vary from species to species and within a species during good health Variation compatible with good health is relatively narrow

Classically proteins were divided into heterogeneous fractions—*albumin* and *globulins* In the development of knowledge on plasma proteins, the nomenclature has varied tremendously being based on methods of detection, fractionation and chemical properties The tendency now is to name plasma proteins on the basis of their function once they have been isolated and characterized, even though an empirical nomenclature continues to exist There seems to be a specific function for each protein that is separated and characterized The major functions of plasma proteins include the maintenance of colloid osmotic pressure, pH and electrolyte balance, and the transport of metal ions, fatty acids, steroids, hormones, and drugs The proteins are in part available as a nutritional source of amino acids for tissues There is an entire family of proteins related to hemostasis and the prevention of intravascular thrombosis There is another

system of proteins that is related to defense against infection The proteins with coenzymatic function exist in inactive forms (proenzymes) There are complex systems of activators and inhibitors which turn the systems on and off as required

Plasma proteins are of great interest in the study of evolution and phylogenetic relationships (Engle and Woods, 1960) In studies of the course of the evolution of vertebrates, it has been shown that cyclostomes lack hemoglobin and fail to produce γ-globulins and ceruloplasmin (Papermaster et al , 1962) Immunochemical procedures are now demonstrating that there may exist a high degree of genetic polymorphism of the plasma proteins in addition to a congenital absence or decrease in several proteins These differences are used to great advantage by human geneticists in study of races and their inherited traits related to disease

The biosynthesis and metabolism of the various serum proteins is beyond the scope of this chapter These topics have been reviewed by Anker (1960) The techniques of separation, identification, purification, and characterization of plasma proteins have progressed tremendously From the traditional salting out techniques, a whole series of chemical and physical techniques has been added, including electrophoresis, ultracentrifugation, combination of cold, ethanol, pH and ionic concentrations, paper electrophoresis, gel and column electrophoresis in addition to new techniques that are being developed regularly Details are available (Neurath, 1965)

Characteristics and Function of Important Plasma Proteins. A series of techniques has become exceptionally valuable in clinical medicine to characterize the normal pattern and variations in disease These techniques are *1*) the Tiselius or free boundary electrophoresis, *2*) paper electrophoresis, *3*) starch gel electrophoresis, and *4*) immunoelectrophoresis These are illustrated in Figure 4 1

The amino acid composition of plasma protein varies considerably (Putnam, 1960) The 18 common amino acids plus the 3 derivatives asparagine, glutamine, and cystine appear to be present in all proteins. The amino acid analysis and sequence determines primary structure The secondary, tertiary, and quaternary structures are responsible for function such as transport (O_2, Fe, Cu, etc.) and enzymatic activity Disease is due to abnormal structure, e g , in sickle cell anemia

There is considerable variation in the end groups and terminal sequences, all of which bear on structure, conformation, and the reactivity of the protein and its function In addition, proteins contain substances other than amino acids, which will be discussed in more detail under the specific plasma proteins

Serum Albumin. Albumin is in the greatest concentration, its obvious major functions are osmotic pressure control and transport Although albumin is considered to be a single homogeneous protein there is a rare hereditary condition in man called bisalbuminemia in which two electrophoretically different forms are observed (Gitlin and Janeway, 1960) There is some variation in the estimate of the molecular weight of serum albumin An average value is 65,000

Albumin has a high content of basic and acidic groups and very little tryptophane Albumin has a free sulfhydryl group which combines with cysteine or glutathione by a disulfide linkage The reactive sulfhydryl group explains the capability of albumin to form dimers with divalent mercury or organic mercurials This reactive sulfhydryl plays an important role in governing the properties of serum albumin

Albumin has a very strong affinity for anions and many other substances with which it can interact Albumin is the major plasma protein involved in the blood transport of anions, cations, dyes, drugs (antibiotics, barbiturates, digitonin, salicyate, sulfonamides etc), thyroxine, and some enzymes such as cholinesterase

Glycoproteins (Mucoproteins). These proteins contain carbohydrate covalently bound Most contain hexose, hexosamine, and sialic acid but little uronic acid Glycoproteins are distributed throughout all of the globulin fractions Glycoproteins have a wide range in molecular weight and carbohydrate concentration and some of the specific ones will be discussed separately Further details on these proteins are available (Neurath, 1965)

Lipoproteins. These proteins are of considerable importance in lipid transport and metabolism They appear to be intimately related to arterial vascular disease They are separated by ultracentrifugal flotation methods They are easily denatured They are classified on the basis of their ultracentrifugal separation into

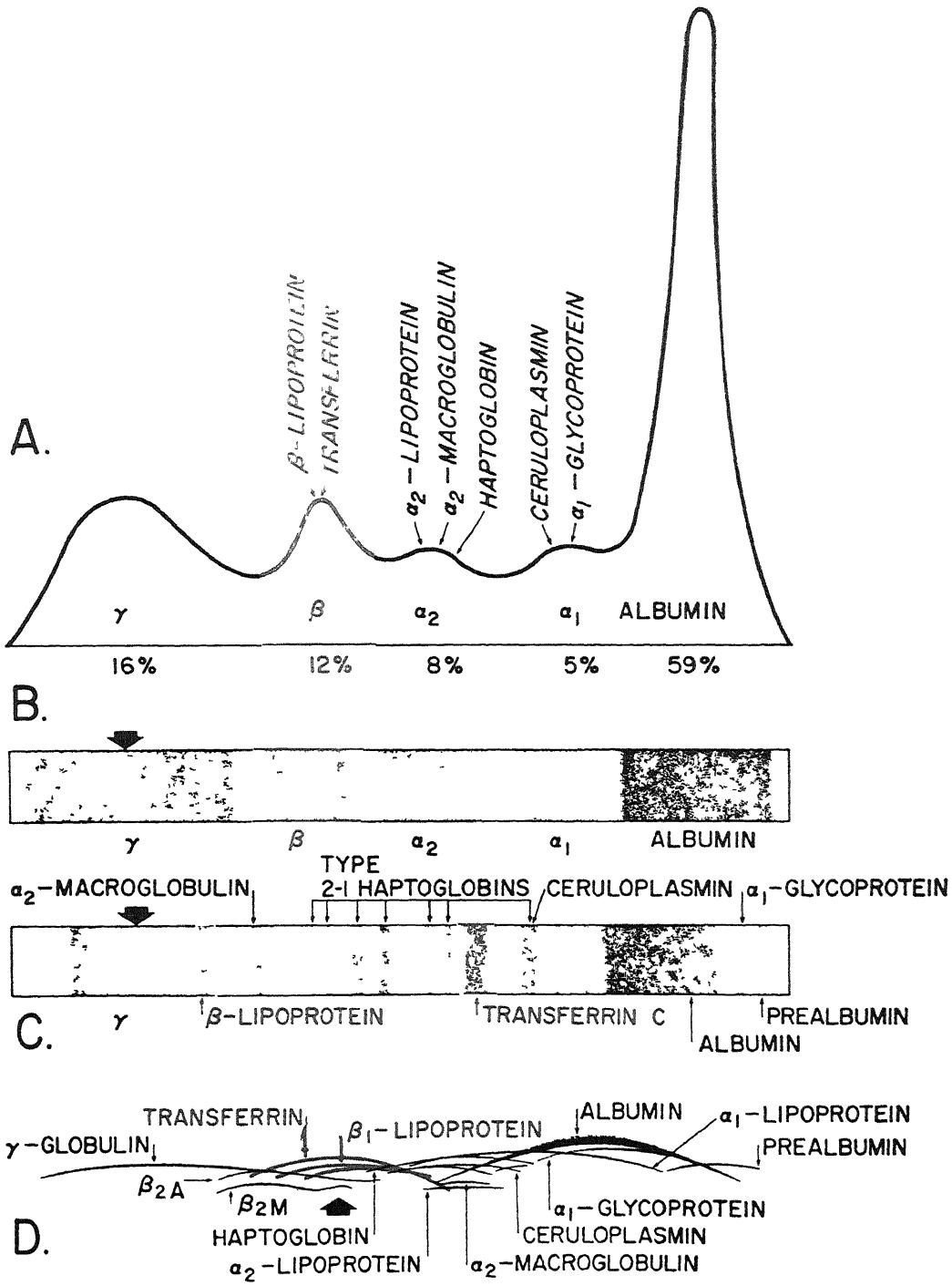


Fig 4 1 Schematic representation of the electrophoretic patterns of normal human serum in pH 8 6 buffer as obtained by four methods *A*) Tiselius or free boundary electrophoresis, *B*) paper electrophoresis, *C*) starch-gel electrophoresis, and *D*) immunoelectrophoresis The broad vertical arrow indicates the starting point in each case β_{2M}-Globulin remains in the starting slot in starch-gel electrophoresis but moves in the γ- to β-range in other methods

low density lipoproteins and high density lipoproteins The low density fractions contain a large amount of glycerides As the density increases, glycerides decrease, and cholesterol esters and phospholipid increase In cardiovascular disease, obesity and liver disease, these proteins are of growing importance

Transferrin. Transferrin is a glycoprotein, molecular weight of roughly 90,000 It has the specific property of iron binding, in which each transferrin molecule combines with two atoms of ferric ion The combining is reversible It transports iron from the gastrointestinal tract to the bone marrow and tissue storage organs Transferrin is intimately related to the regulation and control of iron absorption Its concentration is the limiting factor of iron absorption, thus playing a role in protection against iron intoxication Transferrin is colorless in the absence of iron and pink in its presence It has absorption maximum at 370 mμ and minimum at 410 mμ

There are numerous genetic variations of human serum transferrin The genetically determined polymorphism of human and primate transferrins has been reviewed by Giblett (1962) There are 14 different molecular species of human transferrin characterized by starch-gel electrophoresis under specific conditions The phenotype transferrin C has a high frequency in all populations There are variants in the Chinese, American Indian tribes, Caucasians, Negroes, and other racial groups

Haptoglobins. The haptoglobins constitute a family of glycoproteins which are found in the α-2 globulin fraction of many mammalian species and which have the capability of forming specific stable complexes with hemoglobin (Jayle and Boussier, 1954) Haptoglobin has two very important functions First, haptoblobin binds free hemoglobin, thus preventing loss of iron through urinary excretion It is also believed to protect the kidney from damage by hemoglobin, and regulates the renal threshold for hemoglobin The clinical importance is described in Chapters 4 to 6

There are also marked genetic polymorphism and racial differences in the haptoglobins (Giblett, 1962)

Ceruloplasmin. This copper-binding protein with oxidase activity was named ceruloplasmin because of its blue color It is a glycoprotein present in the α-2 globulin fraction that firmly binds more than 90% of the serum copper The crystalline protein has a copper content of 0 34% with about 8 atoms of copper per mole of protein (molecular weight 150,000 g) The chemical and physical properties of ceruloplasmin have been described by Deutsch et al (1962)

In Wilson's disease (progressive hepatolenticular degeneration), there is a marked deficiency of ceruloplasmin The condition is inherited as an autosomal recessive trait The frequency of the abnormal gene is very low, it is usually the result of consanguineous marriages The total plasma copper is remarkedly reduced, the ceruloplasmin concentration is very low and there is very high urinary copper excretion along with a high concentration in the liver

Fetuin. Fetuin is a protein found in the fetuses and the newborn, first identified in the calf This protein has a molecular weight in the vicinity of 45,000 It has a high sialic acid content It is of great interest in cell proliferation because it is a growth promoting requirement for mammalian cells in tissue culture

Coagulation Factors. These will be discussed in a Chapter 8

The Immunoglobulins These will be discussed in Chapter 10

VARIATIONS IN PLASMA PROTEINS

The protein fractions of plasma change in value independently of one another with or without alterations in the quantity of total protein In some diseases the albumin and globulin fractions may change in opposite directions In hemorrhage there is a loss of all of the fractions of plasma protein The initial restoration of blood volume is by the passage of a solution of low protein from tissue spaces into the blood stream, hence, the concentration of all plasma proteins will be diminished In burns and various diseases of the gastrointestinal tract, the total plasma protein content is reduced as a result of a leakage of blood and plasma from the denuded surface of the skin and into the tissues in the region of the burned area There may be greater loss of fluid than protein, and one may find an elevation in the protein concentration of the plasma Such also occurs in severe dehydration as a result of deprivation of water, or loss of water following excessive urine production as in diabetes insipidus

There may be a reduction of serum albumin in severe malnutrition chronic liver disease cirrhosis of the liver exudative diseases of the gastrointestinal tract, and in severe nephrotic syndrome in which there is an excessive loss of protein via the kidney As the plasma oncotic pressure is reduced because of a decrease in albumin concentration, there is a concomitant loss of fluid into the tissues which may produce a decreased plasma volume Physiologically, there is a low plasma albumin in pregnancy and in the newborn

Plasma fibrinogen is made solely in the liver Hypofibrinogenemia is a rare congenital disease In certain diseases such as carcinoma of the prostate, and occasionally following extensive pulmonary and cardiac surgery, there is a marked decrease in fibrinogen concentration associated with an increased fibrinolysis This is most marked in disseminated intravascular coagulation which will be discussed in Chapter 8 Fibrinogen is increased in acute infections, malaria, tissue injury chronic infection, and physiologically in pregnancy and menstruation Whenever there is an increase in fibrinogen or one introduces an asymmetric molecule into the blood (such as high molecular weight dextran), there is an increased erythrocyte sedimentation rate

The γ-globulin shows an absolute increase in several diseases such as multiple myeloma, cirrhosis of the liver, subacute yellow atrophy of the liver, acute hepatitis and nephritis, systemic lupus erythematosis, and in some acute and chronic infections

THE BIOSYNTHESIS OF PLASMA PROTEINS

Early in the embryo, plasma proteins are synthesized, the albumin fraction first As the blood vessels form, the embryonic plasma floats the primitive blood cells In the adult, it is established that the liver is the primary site of the production of plasma albumin and fibrinogen also is almost exclusively produced in the liver Their concentration in plasma is reduced following injury to the liver The lymphoreticular system, and in particular the plasma cells, synthesize the entire spectrum of immunoglobulins It is easy to demonstrate the synthesis of immunoglobulins by lymphocytes and plasma cells of all degrees of maturity The removal of plasma proteins by plasmapheresis[1] results in a decrease in their concentration Upon cessation of the procedure the concentration increases quickly towards normal, suggesting a tissue reserve pool that can be mobilized quickly This is followed by a slower increase to normal showing that an increased rate of synthesis has been induced These and other studies confirm that there is a dynamic state of body proteins with an equilibrium between production and destruction (Borsook and Keighley 1935)

The general concept that there is a dynamic state of body constituents has been experimentally demonstrated (Schoenheimer 1942) In fact all biological molecules with the probable exception of DNA in the nondividing cells (brain muscle, cells etc) are in a steady state between synthesis and degradation The measurement of the turnover rates of proteins and other biological substances is made possible by the labeling of precursors with stable or radioactive nuclides When a carbon-14 or nitrogen-15 labeled amino acid is introduced into an animal, its incorporation rate into diverse proteins can be measured By the transfusion of labeled pure proteins into intact animals one can measure the degradation rates From the mathematical analysis of the data and the dilution of the labeled material into the total body "pools" one can calculate the pool sizes and the amounts of material synthesized and degraded per unit time per kilogram of body weight Details of these techniques are available (Putnam, 1960)

Proteins have a specific amino acid sequence and structure that is genetically controlled In the normal steady state, messenger RNA is being formed at a steady rate on the genomes that carry the code for the specific protein The messenger RNA flows into the cytoplasm where its message is read by the ribosomes and the required amino acids are activated and brought to the ribosome by transfer RNA where each is incorporated into the polypeptide in proper order When there is a need for accelerated protein synthesis for any reason, it can be accomplished in principle by a longer mean life span of the messenger RNA or an increased rate of messenger RNA synthesis

[1] Whole blood is removed and centrifuged The supernatant plasma is thrown away and the cells returned to the subject This procedure is called plasmapheresis

For details of this process see White, Handler, and Smith (1968)

NONPROTEIN NITROGEN (N P N)

By the term, nonprotein nitrogen, is meant the nitrogen of those substances, e g , urea, uric acid, creatinine, etc , listed in Table 4 2 They may be extracted from blood or plasma by treating either of these with a reagent, such as trichloracetic acid, which precipitates the proteins, and then by filtering and determining the nitrogen in the filtrate These substances are in part absorbed with, or derived from the food, and in part are the waste products of tissue catabolism The total N P N of whole blood amounts to from 28 to 40 mg /100 ml It constitutes from 1 to 2% of the total nitrogen of the blood Its concentration in the cells is nearly double that in the plasma The proportions of the different constituents are given in Table 4 2

The term, undetermined nitrogen, embraces the nitrogen of ammonia, purines, and other nonprotein substances of unknown or undefined nature

Pathological Variations. The N P N of blood at any given level represents the balance struck between nitrogenous materials formed in the intermediary metabolism of ingested and tissue protein and the excretion of these products in the urine In renal insufficiency, therefore, the nonprotein nitrogen is elevated, and in certain cases may be 10 times the normal On the other hand, a rise in the N P N occurs in conditions which are associated with excessive tissue catabolism, such as infections, fevers, thyrotoxicosis, starvation, or severe malnutrition It is also increased following hemorrhage into the stomach or upper intestinal tract In the later months of pregnancy the N P N is reduced The reduction has been attributed to the diversion of nitrogen to the growing fetus and the reduction of protein catabolism in the maternal tissues

The chief conditions associated with an elevation of the N P N of the blood are adrenal insufficiency, dehydration, hemorrhage into the gastrointestinal tract, infectious fevers, lobar pneumonia, intestinal obstruction, parathyroid intoxication (in animals), peritonitis, and renal insufficiency

CHOLESTEROL

This sterol is present in serum in the free form and as cholesterol esters It is carried in the β-globulins, which contain almost all the lipoproteins The value for total serum cholesterol exhibits wide ranges in the normal population, it varies from 100 to 250 mg per

Table 4 2

The nitrogen partition in the blood of normal individuals and the distribution of the various nitrogenous constitutents between the cells and serum

	Corpuscles			Plasma			Whole blood		
	Maximum	Minimum	Average	Maximum	Minimum	Average	Maximum	Minimum	Average
	mg /100 ml			*mg /100 ml*			*mg /100 ml*		
Taken from Wu									
Total nonprotein nitrogen	61	39	49	36	20	29			
Urea N	22	12	17	23	13	19			
Amino acid N	11	8	10	8	5	6			
Uric acid	4	1	2	5	2	4			
Creatine	8	4	6	0	0	0			
Creatinine	3	1 6	2 5	1 5	1	1 2			
Undetermined N			19			2 1			
Taken from Berglund									
Total nonprotein nitrogen	55	38	44	30	18	25	39	28	32
Urea N	13	8	10	17	10	12	15	9	12
Amino acid	11	7	8	6	4	5	8	6	6
Undetermined N	34	18	25	12	2	7	18	10	14

Reprinted from Peters and Van Slyke, 1946

100 ml. The esterified cholesterol is about 72% of the total, and this ratio varies within a rather narrow range. Its concentration is increased in several diseases, notably in hypothyroidism and lipid nephrosis and in some cases of atheromatosis. Keys et al., (1956) found that within fairly wide limits (2 to 3 g. of cholesterol weekly) the content of the sterol in the diet exerted little effect upon the total serum cholesterol.

2 / HEMOPOIESIS

A Vital Cell Renewal System

Introduction

You are a pediatrician You have just seen a child with high fever, swollen glands, and a sore throat The parents believe she has a "strep throat" You diagnose acute lymphocytic leukemia This child will die within months to a few years The chemotherapy of leukemia, as well as the leukemia, will suppress normal production of blood cells making the child susceptible to hemorrhage, infection, and anemia Before death, this child will require antibiotics, transfusions of blood, platelets, and granulocytes There may be major problems with fluid and acid base balance To manage judiciously the treatment of this child, you must understand the stem cell, kinetics of hemopoiesis, fluid, and electrolyte balance Leukemia is a disease of the stem cell The family must be helped to protect the child from knowing life is limited You must make it possible, as long as possible, for the child to play and go to school just as a well child To do this, you must be able to impart your knowledge and concepts of the disease and its management to the parents They must partake in the care of their child Without understanding these can not be done Without understanding of what is known, further progress in the treatment and eradication of this dread disease can not be made

---

The term "cell renewal system" has been suggested as a general way of describing all of those body systems in which the cellular elements are continuously being replaced (Leblond and Walker, 1956) The steady-state cell renewal systems include not only the hemopoietic cells, but also cells of the skin, gastrointestinal tract, lining of the pulmonary system, and the testes These systems are to be contrasted with the relatively immortal cells that are not replaced after completion of growth of the organ (central nervous system, cardiac, and skeletal muscle, etc) and those organs and cell systems in which cellular proliferation is triggered by injury (connective tissues, liver, kidney, etc)

Life depends upon hemopoiesis The red cells transport oxygen and carbon dioxide The neutrophilic granulocytes are a first line of defense against invasion by bacteria The platelets play a crucial role in preventing blood from oozing out of the vascular bed and are essential to staunch bleeding after injury The monocytes play a central role in cellular immunity against specific bacterial infections (tuberculosis, brucellosis, etc) The lymphoreticular system (thymus, spleen, lymph nodes and their relation to the bone marrow) controls the synthesis of specific antibodies against bacterial and viral invaders and foreign tissue, and probably provides a surveillance mission to recognize harmful mutant cells (cancer, etc) and initiate the measures to destroy these clones

Our modern environment may expose man to ionizing radiation in industrial accidents, nuclear warfare, and in the course of treatment of cancer The major cause of death following overexposure to ionizing radiation is suppression of blood cell formation with the concomitant increased susceptibility to infection, hemorrhagic tendency, anemia, and a suppressed humoral immunity All of these alone or in combination may be lethal Aside from ionizing radiation there are many industrial intoxicants that can suppress or eradicate all or part of the blood cell production and produce death Regrettably, many of the drugs that are essential in the practice of medicine may suppress the entire marrow production or selective cell production Lastly, there is a whole series of genetic diseases that result in abnormal function and production of red cells, granulocytes, platelets, or lymphocytes Many of these diseases are either markedly disabling or may result in death early in life

A study of blood occupies a central role in

evaluating the hazards of environment and new and old drugs used in the practice of medicine Accordingly, the physician and physiologist are concerned with the origin of the blood cells and the mechanisms by which their orderly delivery from their sites of manufacture to the peripheral blood is achieved and regulated It is assumed that the reader is familiar with the names and morphology of the mature cells of the blood and of their immediate histologically-identified precursors

ANATOMICAL SITES OF THE FORMATION OF THE BLOOD CELLS

The bone marrow produces red cells neutrophilic, eosinophilic and basophilic granulocytes, platelets, lymphocytes, and monocytes In late fetal life, blood cell production goes on throughout the bone marrow and also within the liver After birth, this function is limited largely to the bones With growth there is a progressive retreat of blood cell production from the distal ends of the long bones, such that in the adult human being, marrow proliferation usually occurs only in the skull, vertebrae, pelvis sternum, and the proximal ends of the long bones Under conditions of prolonged increased demand as in chronic hemolytic anemias, however, blood formation may also return to all of the long bones, out into the fingers and toes, and may erode the cortical bone weakening it to the point that fracture may occur

The lymphocytic-plasma series of cells is formed within the lymphoreticular system consisting of the thymus, spleen, lymph nodes, tonsils and Peyer's patches, and within the bone marrow The interrelationships of the various segments of the lymphoreticular system will be considered in Chapter 10

THE HEMOPOIETIC SYSTEM AS A CELL RENEWAL SYSTEM

The presence of mitosis proves the production of cells within the hemopoietic tissues In the case of the red cell, the average life span has been determined with accuracy to be about 120 days The minimum birth rate of cells can be expressed by the following equation

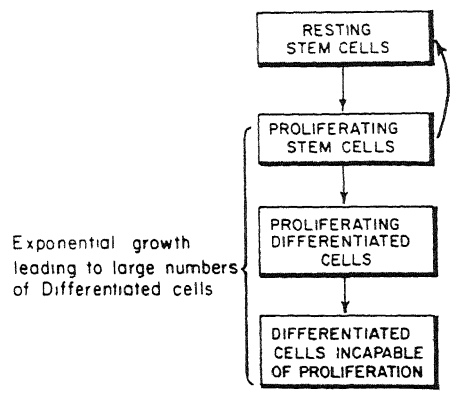
$$\text{Minimum birth rate} = \frac{\text{Total number of red cells}}{\text{Mean life span}}$$

It will be increased if there is any loss in the production line Details on the production will be outlined in the section on erythropoiesis

A Generalized Concept of Hemopoiesis

The functional cells in the peripheral blood contain molecules granules and structures which permit them to perform specific functions These functional cells are the descendants of progenitor cells within the bone marrow which do not contain the molecules or structures characteristic of the mature cell The process by which cells change and develop the capacity to carry out specific functions is called differentiation

The next important difference between mature cells in the blood and their progenitors in the bone marrow is that the mature cells are incapable of mitosis while the progenitor cells divide several times The process of blood formation consists therefore of two essential ingredients *1)* cell *proliferation* and 2) cell *differentiation* In Figure 4 2, a generalized model for the origin of blood cells is shown As the process of cell differentiation proceeds, the



MODEL OF THE ORIGIN OF BLOOD CELLS

Fig 4 2 This is a schematic diagram of the relationship between stem cells and their differentiated descendants Each box represents a population of cells At the top of the figure is a population of stem cells resting either in the G_0 state or in state of prolonged cell cycle Members of this population may pass into a different physiological state characterized by rapid proliferation Cells derived from this proliferation may either retain stem cell properties and return to the resting compartment or may differentiate Many of the cells in the differentiated compartment can divide several times but finally a state is reached in which cell division is no longer possible Cells in the proliferating compartments are growing exponentially so that the differentiated cells greatly outnumber the stem cells

capacity for cellular proliferation decreases The most primitive progenitors have extensive capacity for proliferation and are able to give rise to a large number of differentiated descendants These differentiated descendants have the capacity to fulfill specific functions but have lost the ability to renew themselves Between these two extremes are found the majority of the nucleated cells in the bone marrow These are the cells which have started on the course of differentiation, contain some of the features which will finally define them as functional elements, and are capable of a limited number of cell divisions These are likewise the cells the clinical hematologist recognizes in blood smears obtained from bone marrow, and it is upon their numbers and characteristics that most clinical diagnoses are made

The progenitor cells with almost unlimited capacity to proliferate and also to produce differentiated descendants are called "stem" cells The identity of these cells is subject to controversy Since the stem cells are not yet differentiated, they do not contain the specific cytological features by which they might be easily recognized Stem cells constitute a very small fraction of the cells of hemopoietic tissues Because they have extensive capacity for proliferation, they are greatly outnumbered by their differentiated descendants It has been estimated in the mouse that about one in every thousand cells of the bone marrow is a stem cell Controversy over the stem cell has centered around not only their identity, but also their capacity to give rise to one or more classes of differentiated descendants A vast literature, primarily polemics, based on morphological techniques as to whether there is a single stem cell for hemopoiesis or multiple cells in adult life exist (Bloom, 1938) Today as a result of the elegant studies of the Canadian investigators (Till et al , 1967, Wu et al , 1967, Wu et al , 1968, and Whang et al , 1963), it is known that stem cells with characteristic abnormal karyotypes do differentiate into the erythropoietic, granulocytic, and lymphocytic lines

The identity of the stem cell remains obscure. In the past the existence of pure aplasias of one cell type or another in human disease was used as an argument for the loss of the stem cell for that particular line It seems more likely now that in aplasia either there is a loss of the specific inductor for differentiation of that cell line by induction of the common stem cell, or the stem cell has mutated so it can no longer respond Much is known about the stem cell Radiation protection studies have clearly shown that the common stem cell passes through the blood Yet, the only source for the common stem cell is the bone marrow The stem cell has been described, respectively, as a small, dense mononuclear cell in the bone marrow, and as a small mononucleated lymphocyte-like marrow cell not found in lymph nodes The stem cell is also believed to be a monocytoid cell in the blood (Tyler and Everett, 1966) Aside from the failure to have specific differentiated characteristics, it is probable that the stem cell's appearance changes during its life cycle—proceeding from the resting to the active, proliferating state Since cells double their mass, their size does change To double the mass requires all of the protein synthesizing machinery [ribosomes, messenger ribonucleic acid (mRNA) and transfer ribonucleic acid (tRNA)] to be synthesized and switched on along with the aerobic metabolic machinery (more mitochondria) It seems logical, therefore, that stem cells may have a varying size and appearance as they go through their proliferative cycle

Perhaps more conclusive than arguments from morphology are certain functional tests for the recognition of stem cells Two properties have been described they are the capacity for almost unlimited self perpetuation, and the capacity to give rise to differentiated descendants Without unlimited self-perpetuation of stem cells, marrow function would eventually cease Stem cells must also be able to respond to the normal control mechanisms

An old concept of the stem cell involves what has been called homoplastic and heteroplastic hemopoiesis (Maximow and Bloom, 1948) *Homoplastic hemopoiesis* is the process in the adult where the stem cell for granulocytopoiesis, erythropoiesis, and megakaryocytopoiesis is the youngest member of the series and perpetuates itself, thus providing cells which go through a limited number of mitoses Under physiological conditions requirements are met by homoplastic hemopoiesis During embryonic and later growth, however, or in severe stress or in regeneration from aplasia, the requirements for each cell line are met by

heteroplastic hemopoiesis, a process in which new erythroblasts and myeloblasts are supplied by a pluripotent stem cell (in the parlance of Maximow, a hemocytoblast) This cell is different from the type that in modern terminology is designated by the term, committed stem cell This latter term refers to a cell that responds to stimuli which increase production rate of a specific cell line but is unable to differentiate into all cell lines upon transplantation into fatally irradiated animals The newer concept is that increased erythropoiesis or granulopoiesis due to specific stimuli does not require an input from the pluripotent pool of stem cells, although the latter are responsible for regeneration after aplasias or for regneration upon transplantation into the fatally irradiated animal with aplastic marrow

Before proceeding further with an analysis of cell proliferation in the hemopoietic system, it is necessary to look at the sequence of events that transpires between successive mitoses

THE CELL CYCLE

The cell cycle of mammalian cells is shown in Figure 4 3 Before a cell can divide and increase its number by one, the cell must proceed through a series of stages in which cell organelles, the genetic material [deoxyribonucleic acid (DNA)], enzymes, and structural proteins are doubled The crucial stage in the cell proliferation cycle is DNA synthesis (*S*) during which the chromosomes are replicated precisely so as to transmit all of the genetic information of the species to each of the daughter cells Following the *S* period another series of synthetic events takes place in which the mitotic apparatus is synthesized At mitosis, the cells divide into two identical daughters each of which carries the entire genetic information of the species The cells then enter a period called G_1, a metabolically active period during which enzymes, structural proteins, and cell organelles are synthesized [1] Evidence for this synthesis can be shown by the incorporation of labeled amino acids It has been suggested that hemopoietic stem cells have the capacity to exist in a resting physiological state named G_0 Cells are believed to enter the G_0 state after mitosis This suggests that a collection of cells may consist of a number of cells that constitute a resting, nondividing compartment and another number that represent a dividing compartment (Lajtha et al, 1962) Major advances in the study of cell proliferation were the development of radioactive tritiated thymidine (^{3}HTdR), a specific precursor of DNA and its use in study of mammalian cell proliferation (Hughes et al, 1958) With it, the cohort of cells in DNA synthesis can be labeled and studied by a combination of cytological and radioautographic[2] techniques

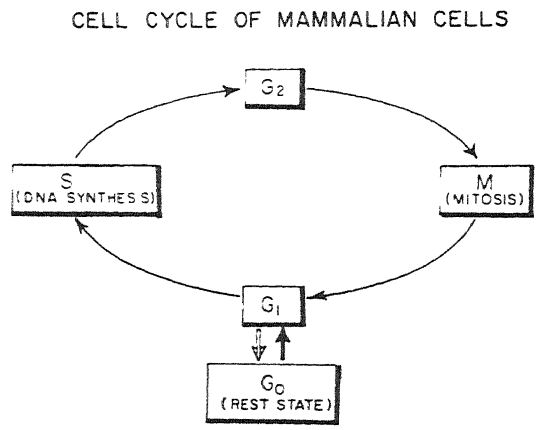


Fig 4 3 A diagrammatic representation of the various phases of the cell cycle

Since one can identify the total population by cytological criteria and the labeled population in S by radioautography, one can define an index of labeling (I_L) as

$I_L = N_s/N$

N_s = Number of labeled cells

N = Total number of this cytological type

It follows that

$I_L = t_s/t_G$

t_s = Average DNA synthesis time and

t_G = Average generation time or the time between sucessive mitoses

[1] The letter *G* stands for gap, in some literature *R* is used representing a rest Since many metabolic activities are in progress throughout cell cycle, the notion of gaps before and after *S* indicating a time period when one sees neither *S* nor mitosis is preferable

[2] Radioautography is a procedure in which a specimen containing a radiation-emitting isotope is placed under a photographic emulsion in such a manner as to permit cells and organs to "take" their own picture The labeled cell is recognized by black reduced silver grains overlying the cell

Similarly one can identify mitosis so that the index of mitosis (I_M) is

$$I_M = N_M/N = t_M/t_G$$

where N_M = Number in mitosis Therefore, the ratio of I_L I_M is

$$I_L/I_M = t_S/t_M = N_S/N_M$$

By serial sampling of cells after the administration of ^{3}HTdR, one can observe the flow of labeled cells through mitosis This is illustrated in Figure 4 4 for a cell cycle without variance and one with variance In the invariant system, for a period of time equal to G_2 no mitotic figures will be labeled As the leading edge of the labeled cohort of cells enters mitosis, the fraction of labeled cells in mitosis will increase, after a time equal to duration of mitosis 100% of the mitoses will be labeled The percentage will remain at 100% until the leading edge of unlabeled cells that were in G_1 enters mitosis, at which time the percentage of labeled mitoses will decrease and after a period of time equal to the mitotic time fall to zero If the various subdivisions of the generative cycle were invariant in a stem cell system there would be repetitive waves as shown in top of Figure 4 4*a* In real life a variance does occur in all biological systems, including in duration of the subdivisions of the generative cycle This produces a rounding out of the ascending and descending curves, and ultimately the fraction of mitoses that are labeled is modulated out to the ratio of the time for the synthesis (t_S) to the time of the generative cycle (t_G) as shown in the middle of Figure 4 4*b* In the bottom of Figure 4 4*c* is shown an actual example of human granulocytic precursors flowing through mitosis From an an analysis of these curves one can in principle estimate the duration of G_2, M, S, and the generative cycle The ascending curve represents the range in the sum of $G_2 + M$ The time between the 50% point on the ascending curve and the descending curve approximates the average DNA synthesis time if the variation in time conforms to a Gaussian distribution

As discussed earlier, there is a reason to believe that the stem cells can be divided into G_0, a nondividing class, and the cells in cycle This situation is illustrated in Figure 4 5 For arithmetic simplicity, half of the population is considered proliferating (D) and half nonproliferating (ND) The index of labeling of the entire population will be equal to the area in synthesis divided by the area of the nondividing

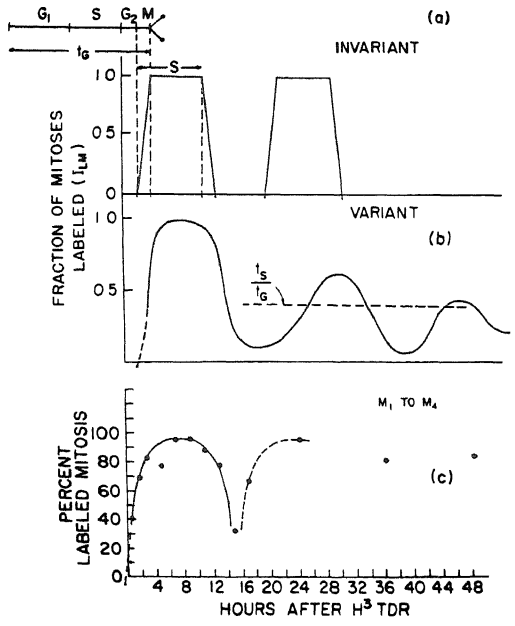


Fig 4 4 *1)* a schematic illustration of the flow of labeled cells through mitosis showing changes in the fraction of labeled mitosis in a self-perpetuating cell system in which there is no variation to phases in the cell cycle *2)* the introduction of biological variability introduces a rounding out of the curves with an ultimate modulation to the ratio of the DNA synthesis time to generative cycle time, *3)* shows the flow of labeled myelocytes of human bone marrow through mitosis from which an average DNA synthesis time of 12 hr is deduced

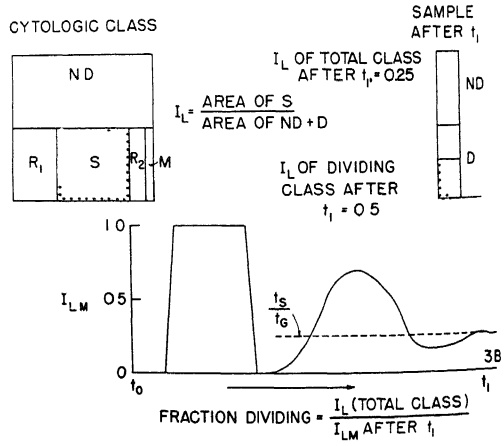


Fig 4 5 Hypothetical cytological compartment consisting of a combination of proliferating cells and non proliferating cells in transit through the compartment showing the means of determining the fraction of cells that is proliferating

plus the dividing populations If one looks only at the cohort of labeled cells moving through mitosis after the minimum duration of $G_2 + M$ the fraction of labeled mitotic figures will rise promptly to 100% and remain there until the unlabeled cells that were in S arrive Then after an interval equal to M the fraction will fall to zero In a system with variance in each subdivision of the generative cycle the fraction of mitotic figures that are labeled would ultimately become modulated to a constant value equal to the ratio of the DNA synthesis time (t_s) to the generative time (t_c) after an interval (t_1), which is the time it takes for the index of labeling to become a constant At this time the I_L of the dividing class under the conditions established would equal 0.5 Arithmetically it can be shown that the fraction of cells that is dividing is equal to the I_L of the total class divided by the fraction of the mitoses labeled (I_{LM}) after t_1 Therefore one can calculate the fraction of cells that are in G_0 (Mendelsohn, 1963, Cronkite, 1964a) Another technique to determine the fraction of cells in cycle is the radioactive thymidine "suicide" technique (Bruce and Meeker 1965) Enough radioactive thymidine is given to kill those cells in S that are incorporating the ^{3}HTdR This technique has shown that about 15% of stem cells in the mouse are in DNA synthesis However, upon transplantation of the surviving G_0 cells into a fatally irradiated mouse, the G_0 stem cells rapidly enter cycle proliferate, differentiate, and restore the bone marrow

CONCEPT OF PROLIFERATION IN THE STEM CELL POOL

The stem cell pool has been visualized as a cytological compartment with autonomous proliferation in which the dividing population doubles with each generation so that the following equation describes the growth (Johnson, 1961)

$$N = N_0\, 2^{t/t_c} = N_0\, (e^{0.693})^{t/t_c}$$

and the growth constant (λ)

$$\lambda = 0.693/t_G \text{ or}$$
$$N = N_0\, e^{\lambda t}$$

where

t_G = generation time
t = time from t_0 to t_1
n = the number of generations or mitoses (n)
λN = λN

the growth rate in an expanding population or the number of cells produced per unit time

If from death or differentiation the rate of removal from the expanding population is $\lambda_1 N$ then the net rate of change is

$$dN/dt = \lambda N - \lambda_1 N$$

and in the steady state is zero

If there is a demand for more stem cells to be differentiated into functional lines to compensate for hemorrhage or a need for more granulocytes to fight infection the demand can be met by a reduction in t_c. This increases λ and t/t_c, the number of mitoses (n) which may take place in the same interval (t) and thus multiplies the output of stem cells The same result can be accomplished by a shifting of more G_0 cells into cycle thus increasing N and with same t_c the dN/dt increases In either case there will be more cells in cycle and the number of stem cells susceptible to agents that kill in S will be increased Such has been observed when the marrow is stressed or when marrow is transplanted into an animal that has an aplastic bone marrow from prior irradiation

If there is not an increased rate of differentiation or death this stem cell compartment will increase in size Something comparable to the latter may be occurring in leukemia

If $\lambda_1 N$ exceeds λN and the stem cell compartment does not sense the need for more stem cells the system will soon become exhausted—a situation which may resemble some hypoplastic marrow states

In this mathematical model one can visualize considerable flexibility, providing that the stem cell pool is "size conscious" Following an unexpected drain of cells by differentiation one can imagine that the remainder may not be susceptible to differentiation until the size is restored by proliferation and restoration of the G_0 population This implies that differentiation probably takes place from the G_0 population The temporary restriction of no differentiation with the same or shorter generation time will permit the G_0 fraction of the stem cell pool rapidly to rebound In fact, one might conceive of its rebounding to larger than the normal size before it is modulated back to preperturbation size In fact, such has been observed experimentally by Lajtha et al (1969)

PROLIFERATING POOL OF DIFFERENTIATED CELLS

The flow of cells is illustrated schematically in Figure 4 2 The structure of a proliferating differentiated compartment is shown in more detail in Figure 4 6 The main characteristic that is imparted to a differentiated pool is multiplication but inability for self perpetuation If the input fails, the compartment vanishes as maturation continues (The compartment is defined in terms of cytoplasmic and nuclear structure) Cells may divide within the cytological compartment one or more times, or may be in transit through it while maturing The mitosis that is seen in some cells shows that division does occur within the compartment Since "flash labeling" by ^{3}HTdR is observed, some cells spend at least some of the time within the cytological phase in DNA synthesis There is no way to determine if all cells that enter a cytological phase divide within this phase one or more times, or if they divide at all Conceptual and mathematical problems concerned with compartment and intracompartmental growth are discussed in detail elsewhere (Johnson, 1961, Killmann et al , 1962, and Quastler and Sherman, 1959)

Within a cytological compartment, the boundaries of which can be determined by morphological criteria, one can observe the fraction of cells that are in mitosis and the fraction of cells in DNA synthesis If the boundary of a cytological compartment is terminated by mitosis, the following simple relations pertain as shown earlier in this chapter

$$I_M = t_M/t_G \qquad I_L = t_S/t_G$$

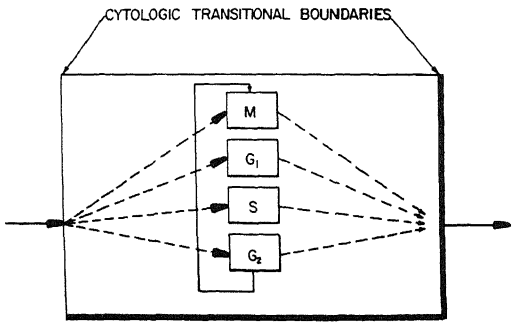
and

$$I_M/I_L = t_M/t_S$$

If one can then determine in a system the index of mitosis and labeling and synthesis time, one can compute the time for mitosis and then substitute in the appropriate equation for calculation of the time for generation

In cell compartments where mitosis occurs between the cytological boundaries, the above simple relationships do not hold For a detailed discussion see Killmann et al , (1963)

In steady-state catenated, proliferating cellular systems, one can determine the relative production rate of cells in each successive cytological compartment normalized to the relative abundance of the most immature definable compartment The first step is the determination of the proportional distribution of cells from the most immature to the most mature in the cell compartment by counting and identifying enough bone marrow cells Next, from the "flash" labeling and mitotic index of each cytological compartment, one can determine the relative number of cells in DNA synthesis or mitosis In Figure 4 6, K_B, the flux of cells from a preceding compartment and the flux out of the compartment under consideration are represented by K_{in} and K_{out} K_b of the intracompartmental birthrate by mitotic division is equivalent to the number of cells in DNA synthesis (N_s), which in turn is equal to the fraction of cells labeled (I_L) times the relative number in the compartment (N) divided by t_S Thus, the outflow from the compartment (K_{out}) equals the birthrate within the compartment (K_B) plus the influx into the compartment, assuming that no death occurs within this compartment In these computations no assumption is made as to whether a cell dies within the compartment or leaves the compartment in some subdivision of the cycle An essential datum for these calculations is the average time for DNA synthesis The method of obtaining this was shown earlier in the discussion in Figure 4 4, the average DNA synthesis time is approximated by the time between the 50% labeling of mitotic figures on the ascending curve and 50% labeling



Fine Structure of a Proliferating Compartment

Fig 4 6 Schematic presentation of a proliferating cytological compartment within the differentiated proliferating pool There is no capacity for self-perpetuation only multiplication Entering cells may cycle one or more times or proceed through only a part of the cell cycle depending upon the rate of maturation or time before they resemble the next cytological class (From Quastler and Sherman, 1959)

on the descending curve of labeled mitosis In Figure 4c, this time is about 12 hr for DNA synthesis in the granulocytic precursors of human bone marrow

The next essential datum is the rate of replacement of the nondividing compartment by labeled cells flowing from the last dividing compartment This was shown in principle in figure 4 6 From the slope of the increasing labeling index of the nondividing compartment, one can estimate the average turnover time of this compartment

The following are clear for any compartment between the stem cell and the first nonproliferating compartment

$$K_{out} = K_{in} + [N_{m \text{ or } s}/t_{M \text{ or } S}]$$

Where K_{out} equals the rate of leaving the compartment, K_{in} equals the rate of entry, $N_{m \text{ or } s}$ equals the number in mitosis or synthesis Therefore,

N_m/t_M or N_s/t_s
= the intracompartmental growth rate (K_B)

The entire proliferating pool as shown in Figure 4 2 consists of a series of cytological phases If the stem cell input K_{in} is known, then the efflux K_{out} from the terminal dividing cytological phase can be expressed as follows

$$K_{out} = 2^n K_{in} = 2t/t_G K_{in}$$

Where t equals the sum of the cytological phase transit times, t_G equals generation time, and n equals the average number of mitoses of t/t_G

FLOW THROUGH A NONDIVIDING STORAGE COMPARTMENT

The rate of replacement of nondividing cells by labeled cells flowing from the proliferative compartment is a measure of the replacement rate A schematic example is shown in Figure 4 7 As the labeled cells move into the nondividing cytologically identifiable compartment the I_L increases From the slope of this curve one can calculate the turnover time[3] of the nondividing compartment If unlabeled cells

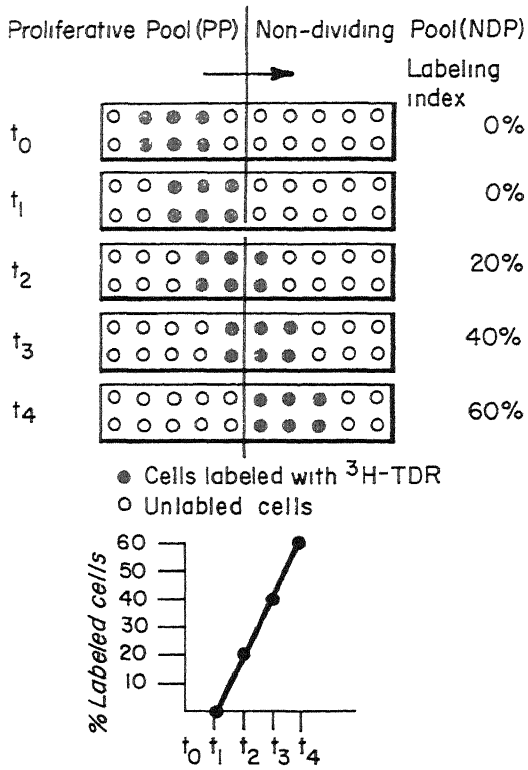


Fig 4 7 Schematic representation of the transition of cells from a proliferating pool (PP) to an initially unlabeled nondividing pool (NDP) Open circles unlabeled cells Filled circles labeled cells 7 *a* From a certain time only labeled cells enter the subsequent compartment It is assumed that the pools have identical turnover times (From Messner, et al 1969)

were to make the transit with labeled cells the turnover time would be less and turnover rate greater than estimated It is also evident that the turnover time for constant input rate is directly related to the size of the compartment that is being replaced (Cronkite et al., 1960a, Messner et al , 1969) For example, if 100 cells/hr enter a compartment containing 100 cells the turnover time is 1 hr —if the compartment has 500 cells the turnover time is 5 hr The reciprocal of the turnover time is the turnover rate in percentage or fraction per unit time, e g , 100%/hr or 20%/hr In the case of successive compartments of varying size flowing from one to another, the average turnover time is proportional to the ratio of the sizes For example if three compartments, A, B, and C, vary in the ratio of 1·2:3 and the turnover time of A is 10 hr., the turnover times of B and C are 20 and 30 hr., respectively.

[3] Turnover time is the time it takes to replace a number of cells equal to the number in the compartment at time zero This does not mean that every cell in the compartment at time zero is replaced, there may be stragglers Nevertheless, a number equal to those present are replaced

LIFE SPAN AND TURNOVER STUDIES IN A FUNCTIONAL NONDIVIDING COMPARTMENT

The classical method in hemopoiesis is to transfuse a known number of immunologically identifiable red cells into the blood, and by immunological differential agglutination techniques determine the time it takes for the transfused cells to disappear By this technique Ashby (1919) established the mean life of the human red cell to be 120 days Several isotopic techniques have been developed to study life span of red cells, granulocytes, lymphocytes, and platelets and will be discussed in Chapters, 3, 7, 8, and 10

Methods of Study of Hemopoietic Stem Cells

It has long been known that one can transplant genetically identical bone marrow into an otherwise fatally irradiated animal whose bone marrow has been eradicated by large doses of ionizing radiation In these hosts, the transplanted cells proliferate and differentiate, providing the recipient animals with the functional cells necessary for their survival from radiation damage Marrow transplantation in mice has been used to obtain populations of hemopoietic cells, each derived from a single cell (Till and McCulloch, 1961) By analysing such clonal populations it is possible to examine the properties of the cells of origin, and in this way to determine the extent to which such cells may be considered hemopoietic stem cells (Till et al , 1964) The technique used to obtain clones is as follows Heavily irradiated animals are injected with genetically identical bone marrow cells derived from healthy donor mice. After 9 to 14 days, the recipient animals are killed, and their spleens are found to contain discrete nodules, which, on histological examination prove to be colonies of proliferating and differentiating hemopoietic cells

Evidence that spleen colonies are clones was obtained by the use of the following experimental techniques (Becker et al , 1963) Recognizable and unique chromosomal abnormalities (markers) were induced by irradiation in some of the cells used to initiate the growth of colonies. When the colonies had developed, their dividing cells were examined for markers In any colony where a particular marker was present, this same marker was found in all of the dividing cells of that colony This provided excellent evidence that all of the cells of such a colony were derived from a single progenitor possessing this marker Enough colonies with markers were found to establish the proba bility that all or almost all spleen colonies are clones

Analysis of clonal populations obtained by the spleen colony technique provides evidence that the cells from which the colonies grow are indeed stem cells First, after 10 days growth, spleen colonies contain more than a million cells, indicating their origin from cells with extensive proliferative capacity Second, spleen colonies have been shown to contain cells themselves capable of colony-formation, indicating that self-renewal has occurred Third, many spleen colonies contain recogniz able erythroblasts, granulocytes, and megakar yocytes, showing that the original cells could differentiate along these three lines Finally, the colony-forming efficiency of cells from the marrow and spleen of normal mice remains relatively constant during adult life, indicating that cells capable of colony formation respond to homeostatic control mechanisms Thus, cells with colony-forming potential have the properties of stem cells, and it appears justifi able to refer to them as hemopoietic colony-forming stem cells

Other assays for stem cells such as radioprotection, erythropoietic repopulation, granulocyte response, marrow cellularity, erythropoietin assay, autorepopulation, and in vitro colony assay are useful, and have been reviewed by Lajtha (1970) and with his associates, (1969)

CHROMOSOMAL EVIDENCE FOR A COMMON STEM CELL FOR ERYTHROPOIESIS GRANULOCYTOPOIESIS AND PLATELET FORMATION

The evidence given in the preceeding sections shows that hemopoietic colony-forming stem cells have capacity to differentiate along erythrocytic, granulocytic, and megakaryocytic lines Evidence for stem cells in man with similar capacity for differentiation has been obtained from the study of chronic myelogenous leukemia In this condition, there is a great increase in bone marrow cell proliferation with the production of large numbers of granulocytic cells at various stages of development

Recently, techniques have been devised which permit detailed study of the chromosomes of mammalian cells (Nowell and Hungerford 1960). They applied these techniques to the cells of patients with chronic myelogenous leukemia and found a specific abnormal chromosome in almost all these leukemic cells. The abnormal chromosome is usually called the Philadelphia chromosome, after the city in which it was first recognized. Whang et al. (1963) have combined chromosomal analysis with other functional tests, and found that the Philadelphia chromosome is present not only in the cells of the granulocytic series but also in the progenitors of erythrocytes, and certain large polypoid cells, almost certainly megakaryocytes. Because these relatively differentiated progenitors of red cells and platelets have only limited proliferative capacity it was concluded that they are the partially-differentiated descendants of leukemic progenitors. Since the Philadelphia chromosome, a specific marker, is found in granulocytic erythrocytic, and megakaryocytic cells, but not in other dividing cells of the body, a single progenitor must possess the capacity to differentiate in these three different directions. The existence of a pluripotent leukemic cell is strong evidence that there also exists a normal human stem cell with capacity for differentiation towards erythrocytic, granulocytic, and megakaryocytic pathways. These plus earlier discussions suggest that there is a pluripotent stem cell. Committed stem cells will be discussed in Chapters 3 and 7 on erythropoiesis and granulopoiesis, respectively.

There are two clinical situations in which stem cell potential and kinetics are of major importance. In genetically identical animals, it has been shown that bone marrow transplantation will save the life of otherwise fatally irradiated animals. There are numerous drugs in addition to radiation that may produce bone marrow failure. An example is aplastic anemia produced by the toxic effects of an antibiotic, chloramphenicol, which is used in therapy of some bacterial infections. If aplastic anemia develops the patient is in danger of dying because he lacks the class of stem cell which differentiates along erythrocytic, granulocytic, and megakaryocytic lines. From animal experimentation, the physician knows that production of these cells can be re-established by transplantation of genetically identical bone marrow. Yet both animal and clinical studies show that in attempting such a transplantation one also transplants cell with immunological capability that will flourish, proliferate, and grow out cells that recognizing the immunological difference of the host, will launch an immunological attack upon the patient. This disease is called a "graft vs host reaction," and will be discussed in more detail in the section on the lymphoreticular system (chap. 10). If the recipient's immunological system is likewise functioning, or becomes reestablished, it can then recognize and launch an immunological attack against the transplanted stem cells and by rejecting the transplant thus defeat the purpose of the attempted therapy. The trick then is to transplant the common stem cell so as to permit differentiation into erythropoiesis, granulopoiesis and thrombopoiesis, but to suppress simultaneously its differentiation into immunologically competent cells that attack the host. Furthermore the reaction of the host's immunological system against the foreign transplanted cells which will reject the life saving graft must also be suppressed. To attain all of these objectives at the same time is extremely difficult. Nevertheless, major efforts are now underway in attempts to transplant bone marrow and immunologically competent cells as a means of treating otherwise fatal diseases of man.

The other situation in which stem cell potential and kinetics is vital is in the chemotherapy of leukemia. All of the drugs used to treat leukemia are nonspecific and promiscuously injure proliferating normal and leukemic cells. If it were possible to put the normal stem cells into a more resistant phase of cell cycle during treatment there would be a greater probability of eradicating the leukemia without eliminating the stem cells required for life. To date this has not been accomplished. The apparent leukemic cures result from the fact that stem cells for normal hemopoiesis tend to recover more rapidly than the leukemic stem cells (shorter generation time in normal than leukemic stem cells), and thus may establish a relatively normal hemopoiesis for variable periods of time while the leukemic clones of cells are growing out again more slowly. Further understanding of stem cell behavior and kinetics in man is essential for handling of several diseases of major importance.

Certain types of anemia in the mouse are genetically controlled. Detailed studies using the marrow transplantation technique of Till and McCulloch have shown that one inherited anemia is produced through deficient colony formation by stem cells, whereas, another inherited trait is characterized by a situation in which the stem cells are normal but a host factor essential for the growth of its stem cells or normal transplanted stem cells is deficient (McCulloch et al., 1964). It is reasonable to expect that some of the fatal anemias of man will also prove to be genetically determined defects in stem cells or host factors.

3 / THE ERYTHROCYTE

Normal Erythropoiesis

A continual delivery of oxygen to the tissues is essential for life This transport is accomplished by hemoglobin, a respiratory pigment capable of accepting, transporting, and releasing oxygen to the tissues The hemoglobin is enclosed within the red cell Serious diseases are produced by aberrations in the structure of hemoglobin, abnormality in the structure of the red cell envelope in which the hemoglobin is enclosed, or abnormalities in the red cell assembly line Under normal circumstances in adult man, all blood cell production takes place inside of the bone marrow Nutrient arteries lead through the cortex of the bone and into the marrow cavity Within the cavity they branch into distributing arterioles, which in turn give rise to the sinusoidal bed The diffuse, fragile, sinusoidal bed coalesces into a central vein that proceeds to a foramen and exits from the bone The sinuses form a system of vessels running radially from the periphery toward the central longitudinal vein .Hematopoietic tissue lies between the sinuses The fully formed sinusoidal wall is trilaminar and consists of an endothelial lining cell, basement membrane and adventitial cells In areas there may be only the lining endothelial cell Sometimes the wall is bilaminar consisting only of lining cell and basement membrane, or lining cell and adventitial cell These latter cells may be phagocytic They also contain fat droplets and may accumulate large deposits of fat characteristic of fatty marrow By electron microscopy, it is seen that cells proliferate outside of the sinusoids, and megakaryocytes can be seen to discharge platelets through mural apertures In some places large segments of the sinus wall may become attenuated, pierced by apertures Normoblasts, reticulocytes, and granulocytic leukocytes enter the circulation by passing into a sinus; they may pass through existing apertures or create an aperture by pressing into a sinus wall. A segment of the sinus wall may disappear permitting extravascular cells to exude into the circulation It is believed that the sinuses and extrasinusoidal cords form a system wherein portions become vascular or extravascular depending upon the requirements for hematopoiesis The blood flow is regulated through the arterioles and delivery of cells to the blood (Weiss, 1965)

Erythropoiesis is one of the better understood cell renewal systems of man There is a clearly defined flow of cells from the primitive, common stem cell to the mature red cell in the blood (see Fig 4 8) In the past, the accepted concept demanded an undifferentiated stem cell pool which is acted upon by erythropoietin, a specific inductor that directs stem cells irreversibly down the erythropoietic pathway (see below) In recent years, however, evidence has accumulated indicating that there are two functional, if not morphological levels of stem cells (see Fig 4 9) One stem cell is pluripotent (not sensitive to erythropoietin), and one is committed to erythropoiesis and is responsive to erythropoietin This dual notion is similar to the very old concept that there may be homoplastic hemopoiesis, the production of mature cells by young elements of the same type, and also heteroplastic hemopoiesis, when increased demands are met by increased production from the common pluripotent stem cell (Maximow and Bloom, 1948) The committed stem cell has been shown to be predominantly in a proliferative cycle, whereas, the pluripotent stem cell is mainly in a resting stage (G_0) as would have been expected by Maximow and Bloom This newer concept is derived from studies in which agents that destroy cells in deoxyribonucleic acid (DNA) synthesis (hydroxyurea and tritiated thymidine, ^{3}HTdR) were used. Administration of these materials drastically reduced the number of cells capable of responding to erythropoietin, but did not significantly influence the transplantability of the undifferentiated colony forming stem cells and their ability to restore all lines of hematopoiesis (Morse et al., 1969,

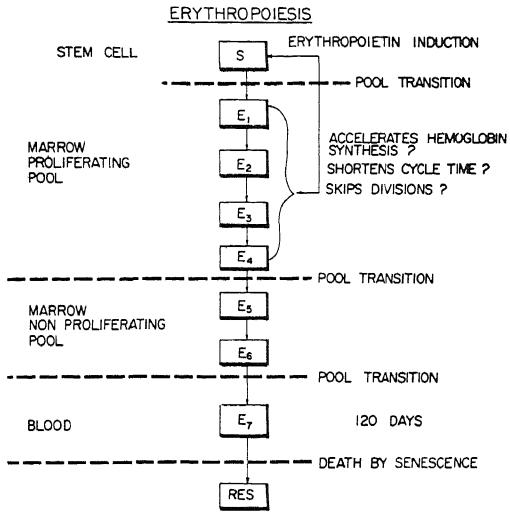


Fig 4 8 Schematic presentation of flow of cells from the common stem cell through the proliferating and non-proliferating cells to ultimate death by senescence Alleged points of action of erythropoietin are indicated (From Cronkite 1964 b)

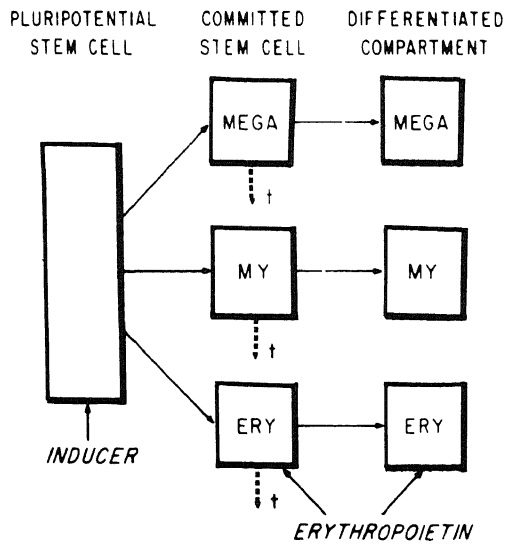


Fig 4 9 A schematic model of hematopoiesis Meg megakaryocytic MY myelocytic ERY erythropoietic Dashed line and cross indicate cell death (From Stohlman 1970)

Iscove et al , 1970) (see chap 2) Although the factors regulating the pluripotent cell are poorly understood, it has been clear that these cells can circulate in the blood stream under stress (Brecher and Cronkite, 1951). It has been suggested that the pluripotent stem cells are in a dynamic equilibrium with tissue stem cells and settle out and grow where there is a favorable microenvironment (Kretchmar et al , 1970) To date, it appears that the pluripotent stem cells are not under humoral control A short range cell-to-cell type of interaction has been postulated as being the method of regulating proliferation and differentiation of pluripotent cells This has been related to the histocompatibility surface antigens and other factors (McCulloch and Till, 1970, Cudkowicz and Stimpfling, 1964) Whether these short range cell interactions are stimulatory or inhibitory remains obscure The concept of the relationship between the pluripotent stem cell and committed stem cells in the differentiated cell lines is shown in Figure 4 9 It is not clear whether dedifferentiation of a committed stem cell back to the pluripotent cell is possible Erythropoietin has been postulated to decrease the cycle time at the expense of the G_1 phase It might be pointed out here that as mechanisms for increasing the cell output, there are three possibilities *1*) increasing the input of stem cells down the differentiated cell pathway with the same number of mitoses during the same period of time, this will increase the production rate in proportion to increased stem cell input and will require additional space, *2*) shortening the generative cycle with more mitoses in the same period of time, this will increase production rate in proportion to the additional number of mitoses, or *3*) a combination of both (See Chapter 2 for discussion of principles involved)

Upon entering the erythropoietic compartment, the stem cell rapidly assumes cytological characteristics of the differentiated line [1] As the cytological characteristics of the nucleus and cytoplasm change, the cells receive successively different names—proerythroblasts (E_1), large basophilic normoblasts (E_2), small basophilic normoblasts (E_3), polychromatophilic normoblasts (E_4), late nondividing normoblasts (E_5), reticulocytes (E_6), and the erythrocyte (E_7) During the stages of E_1 through E_4, one can observe mitosis and "flash" labeling of DNA by ^{3}HTdR in a clearly defined fraction of each cytological class It is generally assumed and/or implied that each cell that enters erythropoiesis divides in each successive stage This has not been proved

[1] See standard hematology atlases for description of cell types

Table 4 3
Relative production rates in erythropoiesis

Normoblastic Hemopoiesis		E_1	E_2	E_3	E_4	E	%/h [a]	Per Cent[b]
Subject I	N[c]	100	217	645	1061	3270		
	I_L[d]	0 80	0 86	0 92	0 55		2 91	
	N_s[e]	80	187	594	584			
	K_{in}[f]	?	6 15	20 55	66 25	95 1		
	K_b[g]	6 15	14 40	45 70	44 90			
	K_{out}[h]	6 15	20 55	66 25	111 15			85 5
Subject II	N	100	220	305	577	1300		
	I_L	0 74	0 78	0 68	0 56		2 91	
	N_s	74	164	208	323			
	K_{in}	?	5 7	18 3	34 3	38 0		
	K_b	5 7	12 6	16 0	24 8			
	K_{out}	5 7	18 3	34 3	59 1			64 5
Subject III	N	100	203	300	616	1410		
	I_L	0 88	0 81	0 65	0 52		2 91	
	N_s	88	165	195	320			
	K_{in}m	?	6 8	19 5	34 5	42 0		
	K_b	6 8	12 7	15 0	26 4			
	K_{out}	6 8	19 5	34 5	59 1			71 1

Messner et al , 1969
[a] Hourly increment of labeled cells appearing in compartment E_5
[b] Percentage of K_{in} E_5/K_{out} E_4
[c] N, relative compartment size
[d] I_L, initial labeling index
[e] N_s, number of cells in DNA synthesis, i e , ^{3}HTdR labeled
[f] K_{in}, hourly rate of entry of cells into a compartment
[g] K_b, hourly birth rate of cells within a compartment
[h] K_{out}, hourly rate of exit of cells out of a compartment

Some cells (as will be discussed later) may jump all the way from an E_1 to an E_4 or E_5 without intervening mitoses

At the level of E_5, mitosis and DNA synthesis are rarely if ever seen The switching off of cell proliferation is thought to be due to the increasing concentration of hemoglobin which, when it attains about 20%, is believed to suppress further cell cycling It is known that an equilibrium exists between the intranuclear and cytoplasmic hemoglobin through nuclear pores It is plausible though not proved that the concentration of this specific protein inhibits further cell proliferation (Stohlman, 1970) Within about 9 hr after the cells cease cycling, the nucleus of the E_5 cell is ejected and the cell becomes a reticulocyte (Bond et al , 1962)

Utilizing the principles described in Chapter 2, the relative production rates of red cell precursors is shown in Table 4 3 In the calculation of birth rates, a DNA synthesis time of 12 hr was used (Stryckmans et al , 1966) The transit rate from E_4 to E_5 was determined by the movement of the labeled cells from the proliferating compartment into the first segment of the nonproliferating compartment (see chap 2, Fig 4 6 for principles) In Table 4 3, the replacement rate of E_5 by labeled E_4 was 2 91%/hr. for a maximum turnover time[2] of 34 hr If unlabeled cells made the transit in parallel, the turnover rate would be faster and turnover time shorter

If the rate of entrance of cells into E_5, the first nondividing compartment, equals the rate of production of cells in E_4, there is no loss of cells by death in the assembly line, if more cells are produced in E_4 than appear in E_5,

[2] Reciprocal of turnover time is turnover rate.

there is a loss of cells These data show that there was an apparent loss between E_4 and E_5 of 15 to 35% (average of 25%) for the three individuals This is a very crude preliminary estimate However, some ineffective erythropoiesis would have been predicted (Finch, 1959)

After loss of the nucleus, the cell is called a reticulocyte because it contains residual ribonucleic acid (RNA) (ribosomes, messenger and transfer) which stains with new methylene blue or brilliant cresyl blue and thus makes it possible to identify morphologically this cell type Reticulocytes mature to red cells in the peripheral blood with an average maturation time of 33 hr in vivo (Stryckmans et al , 1966)

Further insight into the kinetics of red blood cell formation in man can be obtained by looking at the turnover of red cells If one assumes a red cell concentration of 5×10^6/mm ^{3}, an average life span of 120 days and a blood volume of 6 9% of body weight, the turnover rate of red blood cells (RBC) in the blood (RTR) is

$$\text{RTR} = \frac{\text{Red cell per unit volume} \times \text{blood vol /kg body weight}}{\text{RBC life span in days} \times 24} \quad (1)$$

$$= 5 \times 10^6 \times 10^3 \times 0\,060 \times 10^3/120 \times 24$$

$$= 12 \times 10^7 \text{ red cells/kg /hr}$$

Since data discussed earlier suggest a loss of 25% of nucleated red cells in the assembly line, the estimated red cell precursor marrow production rate (RPR) is

$$\text{RPR} = \text{RTR}/0\,75 = 16 \times 10^7/\text{kg /hr}$$

The total number of nucleated red cell precursors has been measured (Donohue et al , 1958), there are $5\,3 \times 10^9$ nucleated red cell precursors per kilogram By pooling data in Table 4 3, one finds that 0 43 of the nucleated red cells are in the proliferating pool (E_1 to E_4) Hence, this pool is $2\,3 \times 10^9$ cells/kg The ratio of dividing erythropoietic precursors to dividing granulopoietic precursors is (2 3 2 6) or 0 88 The granulocyte turnover rate is blood is $6\,8 \times 10^7$/kg /hr. (chap 7) The ratio of these two turnover rates in the blood cells is $12 \times 10^7 - 6\,8 \times 10^7$ or 1.76 This means that for each respective precursor about two red cells are being destroyed in the blood for each granulocyte. Since there is an apparent loss of some cells in red cell pathway more than two are being made. (It has been suggested by Patt and Maloney, (1964) that some are lost in myelopoiesis) The erythropoietic machinery is producing cells faster, whereas, the granulopoietic machinery appears to be throttled back

The average I_L from the proliferating E_4 compartment is about 0 54 The average t_G is roughly 22 hr The sum of t_S, t_{G2} and t_M is 16 (12 + 3 + 1) hr , hence, t_{G1} is about 6 hr

As mentioned earlier, red cell production can be increased by *1*) an increased number of mitoses in the same period of time between the stem cell and E_5, the nondividing compartment, *2*) an increased flux of stem cells into the erythropoietic line, or *3*) a combination of *1*) and *2*) A doubling of stem cell input of each additional mitosis doubles the output and doubles the space required for erythropoiesis within a rigid bony box if transit time remains the same The space can come from encroachment on other cell lines, sinusoids, and yellow fatty marrow If the generation or transit times decrease, space occupied by the erythroid marrow will diminish and so permit an increased stem cell input or increased number of mitoses in the same space with a consequent increase in red cell output Chronic blood loss or chronic hemolysis is compensated by an obvious increase in the size of the erythroid marrow, and red cell turnover rates six times normal have been observed

If erythroid generation time can be shortened at expense of G_1 and if there are roughly four sequential mitoses, there is a possibility, at least in principle, that the marrow could accommodate another mitosis so as to double the erythrocyte output Further increases would require an increased stem cell input and more marrow space

If there were more extensive comparable data on absolute cellularity of the bone marrow and the fraction in DNA synthesis, one could estimate by principles elaborated in Chapter 2 the "steady-state" stem cell influx into erythropoiesis, the average number of mitoses from the stem cell to E_5, and the average transit time from the stem cell to E_5 Preliminary estimates suggest an average of 2 27 mitoses and an input of stem cells of $3\,3 \times 10^7$/kg /hr It is believed that a small fraction of E_1 or E_2 may proceed directly to the nondividing pool (Alpen and Cranmore, 1959), a notion also described as "skipping mitoses," which is believed to be more frequent when

erythropoiesis is stressed (Stohlman, 1970) Their observations and the failure to compute a whole integer for the number of mitoses suggest that cells divide a variable number of times after commencing down the erythropoietic pathway—a notion inconsistent with most diagrams for hemopoiesis which show a mitosis at each stage of hemopoietic development The presence of mitoses or DNA synthesis only demonstrates that cells can divide or commence replication at that stage but does not prove that they do in fact divide (Killmann et al , (1963) An extra mitosis will double output only if each cell divides an additional time If a significant number of cells proceeds from E_1 to E_5 without this or with even fewer divisions the total output will not be doubled

Emphasis has been placed upon the functional study of the stem cell, and upon the measurement of erythropoiesis by labeling cells with ^{3}HTdR while they are in DNA synthesis and observing their subsequent flow as described earlier There are many other ways of examining erythropoiesis For example, most of the iron used in hemoglobin synthesis is incorporated into the cell at the E_1 and E_2 stage If radioactive iron is used as an in vivo label, grain counts can be made by radioautography of cells that are young, and then subsequently of cells in the later stages Each mitotic division decreases the grain count of a labeled cell by one-half Approximately three mitotic divisions with a generation time of about 20 hr occur during the transit from an E_1 to and E_4 cell (Alpen and Cranmore, 1959) However, some cells apparently jump from E_1 to E_5 without further mitosis

The preceding discussion and calculations are based on very few observations and probably have a large error The examples serve to show principles involved and approaches to an understanding of a complex cell renewal system Further observations may change considerably the actual values for the time parameters of the cell cycle and the cell fluxes

REGULATION OF ERYTHROPOIESIS

The *erythron* is a term used to indicate the combined mass of immature red cells within the bone marrow and mature cells in the blood (Castle and Minot, 1936) Under conditions of health, the erythron is maintained quite constant However, it is very responsive to changes in oxygen tension hemorrhage, hemolysis and disease It was predicted by Paul Bert in 1878 that high altitudes or low oxygen tension would demand a compensation in the mass of hemoglobin in order to transport more oxygen This prediction was fulfilled and vividly observed in the striking effects upon the concentration of red cells in human beings living at high altitudes such as in the Peruvian Andes or the Himalayan mountains The first convincing evidence that there was a substance circulating in the blood that stimulated erythropoiesis came from the studies on parabiotic rats in which it was found that a similar degree of normoblastic hyperplasia and increase in the red cell mass occurred when one rat breathed air at atmospheric pressure and one breathed air with a low oxygen tension (Reissmann, 1950) Clinical studies in patients with right to left flow through a patent ductus arteriosis and consequently, with cyanosis and hypoxia of the lower part of the body and normal oxygen tension in the upper part of the body, have shown that the degree of erythropoietic hyperplasia in specimens of the bone marrow from the upper and lower part of the body are identical (Stohlman et al , 1964) These observations made untenable the hypothesis that oxygen alone directly controls red cell production and produced very strong evidence that as a result of hypoxia there is a material produced that circulates in the blood and controls the rate of red cell production

The presence of a nondialyzable substance in the serum of rabbits rendered anemic by bleeding or made hypoxic by low oxygen tension of the inspired air, a substance that causes reticulocyte responses and increases in nucleated red cells in the marrow when injected intravenously to normal homologous recipients, was demonstrated by Erslev (1959) *Erythropoietin* is now recognized as a glycoprotein hormone and to be an important regulator of erythropoiesis in mammals It is detectable in increased amounts in the body fluids of animals subjected to various forms of hypoxia as well as in oxygen deficiency states in man Erythropoietin is involved in accelerating erythropoiesis in the panic states following massive hemorrhage or exposure to severe hypoxia as a result of high altitude, severe hemorrhage, hemolysis, or cardiopulmonary disease In addition, it presumably plays an essential role in the day-to-day replacement of

red cells The evidence for the twofold role is overwhelming Furthermore, hypertransfusion by red cells will almost completely suppress erythropoiesis

The kidney has been established as a principle although not the exclusive site of production of erythropoietin Nephrectomy in animals results in a marked diminution in the production of erythropoietin following the induction of anemia or low oxygen tension of the inspired air The failure to extract erythropoietin from kidneys led investigators to determine whether it was present in the form of a precursor It has now been shown that there is a factor, called renal erythropoietic factor (REF), which is not related chemically to erythropoietin (Gordon and Zanjani, 1970) This factor is capable of generating erythropoietin when incubated in vitro with normal serum It is present in the mitochondrial fraction of kidney, hypotonic extracts of this fraction contains significant quantities of the material It is now believed that the REF is an enzyme that cleaves a part of the serum substrate to produce functional erythropoietin The system appears to be analogous to the renin-angiotension mechanism

Conceptually, the steady state can be described as in Figure 4 10 Starting with the blood compartment which transports oxygen, one can visualize that as the red cell mass diminishes the delivery of oxygen decreases and allows a tissue hypoxia to develop This results in secretion of the precursor of erythropoietin and its activation by the renal erythropoietic factor The active erythropoietin stimulates the committed stem cell compartment to proliferate, possibly decreases the generation time of erythrocytic precursors and also accelerates the synthesis of hemoglobin throughout the proliferating compartment There is a time

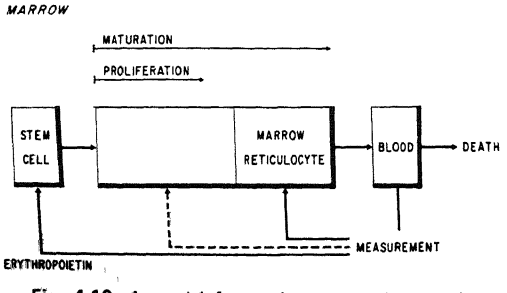


Fig. 4.10. A model for erythropoiesis showing feedback relationship with peripheral blood (From Stohlman 1970)

delay of several days from differentiation through the maturation and proliferation compartments to the reticulocyte and the appearance of the mature red cell in the peripheral blood Whether the disturbance in oxygen delivery has occurred as a result of a decrease in the number of red cells, changes in the oxygen association curve or blood flow, there is a time delay before this disturbance can be corrected It can be argued that such a system, because of the long time delay, will tend to over correct and thus oscillate, the oscillation will have a period of twice the time delay In the case of erythropoiesis, however, because the lifespan of the red blood cell in the peripheral blood is much longer than the stem cell to red cell transit the oscillation will be damped, yet it can be detected by measuring oscillations of the reticulocyte count in the blood (Moreley and Stohlman, 1969) The preceding observations show that erythropoiesis (and, in fact, all hemopoiesis as will be discussed in chap 7) is not a continual steady state in which production at any one moment equals destruction, but rather a controlled state of sustained oscillations in which the production and destruction rates are equal when averaged over an interval equal to the duration of the cycle

Red Cell Structure and Functions

For many years the red cell was considered to be a nonliving cell that transports oxygen It is estimated that in its average 120 day life span it travels 170 miles to deliver oxygen to the tissues The complexity of the red cell was greatly underestimated until relatively recently During its development in the marrow approximately 4×10^8 molecules of hemoglobin are synthesized comprising 95% of its dry weight and 30% of its wet weight The red cells transport about 250 ml of oxygen/min to the tissues in the average man in a basal state It has been estimated that the average blood cell spends about 1 78 sec in a pulmonary capillary and that the oxygenation takes place within about one-third of this time The total surface area of the circulating erythrocytes is 1500 times the body surface (Harris, 1970)

THE SIZE, SHAPE AND STRUCTURE OF THE RED CELL

Human erythrocytes are biconcave, nonnucleated disks having a mean diameter of 7 2

μm (6 to 9 μm, see Fig 4 11) and a thickness of about 2 2 μm (2 to 2 4 μm) at the thickest part i e near the circumference, and about 1 μm at the center As a result of osmotic changes and the consequent passage of water into the cell the diameter increases with a shift in the acid-base balance of the blood toward the acid side The cell is, therefore slightly larger in venous than in arterial blood its diameter is increased by about 0 5 μm in muscular exercise and is decreased by forced breathing The central portion of the cell is much thinner than its edges which appear heaped up into a circumferential mound around a central depression This construction gives it a biconcave contour or a roughly dumbbell outline when viewed edgewise (Fig 4 11) The average area of a red cell is 120 μm and the volume 85 μm

The biconcave disk is uniquely suited for the transport of oxygen A sphere has a surface that is equidistant from the center, and also has the smallest surface area in relation to its mass A thin disk on the other hand presents an almost maximal surface area in relation to its bulk, and diffusion distances are shorter This form secures the advantages of equal and rapid diffusion of oxygen to its interior and a relatively large surface area for the absorption of the gas The biconcave form also gives the blood cell a mechanical advantage, in that the changes in volume which the cell goes through from time to time can be accomplished with a minimal amount of tension being placed upon the cell membrane (Hartridge, 1920)

The red cell consists of 71% water, 28% hemoglobin, 7% lipid (cholesterol, lecithin, and others), and a 3% mixture of sugars, salt, enzyme, protein, and glutathione The ultrastructure of the red cell membrane is very complex The membrane is birefringent indicating a patterned molecular structure Electron microscopy shows the surface of the human red cell to be covered by a layer of circular plaque-like structures approximately 100 × 500 × 30 Å These plaques are believed to be related to channels or pores that allow for the ingress and egress of water and electrolytes The hydrophobic groups of the plaques are projected outward and impart a very strong negative although not necessarily uniform charge to the surface of the cell Beneath the plaque, proteins are arranged in long fibrillar elements called elanin A, B, O and C, D, E antigens are

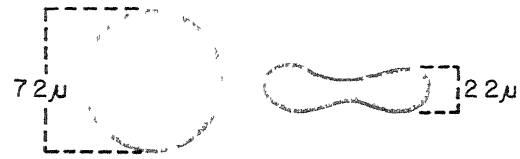


Fig 4 11 Diagram showing dimensions of the red cell

associated with the elanin fractions The membrane is exceptionally well structured at the surface and contains practically no hemoglobin Towards the center there is less order and the concentration of hemoglobin increases Knowledge of the structure of the red cell membrane is the result of 30 years of electron microscopy, where the appearance of membranes at high resolutions is related to the methods used to obtain the pictures The new technique of "freeze cleaving' electron microscopy has permitted considerable progress in study of the membranes of cells Figure 4 12 shows a concept of the red cell structure as revealed by this technique Note that the membrane associated particles may extend through the full thickness of the membrane Although great progress has been made in studying the structure of the membrane chemically and ultrastructurally, the relationship of structure to function is not clearly demonstrated

THE RED CELL AT WORK

An amazing fact is that the red cell does no "work" in fulfilling its primary function of oxygen and carbon dioxide transport Oxygen and carbon dioxide are transported to hemoglobin by diffusion These vital transport functions are passive and do not require energy One may ask then, why does the red cell contain the set of enzymes necessary to carry out anerobic carbohydrate metabolism and the hexosemonophosphate shunt? In addition, it possesses a wide spectrum of enzyme systems capable of directing the utilization of potential energy of the various high energy compounds that are produced This energy is presumably used to maintain the cell as a functioning unit

Figure 4 13 shows the general metabolic pathways of adult erythrocytes Most of these enzymes are phosphate containing The interrelations and the role of these enzymes is shown schematically The red cell retains only two carbohydrate pathways The energy for

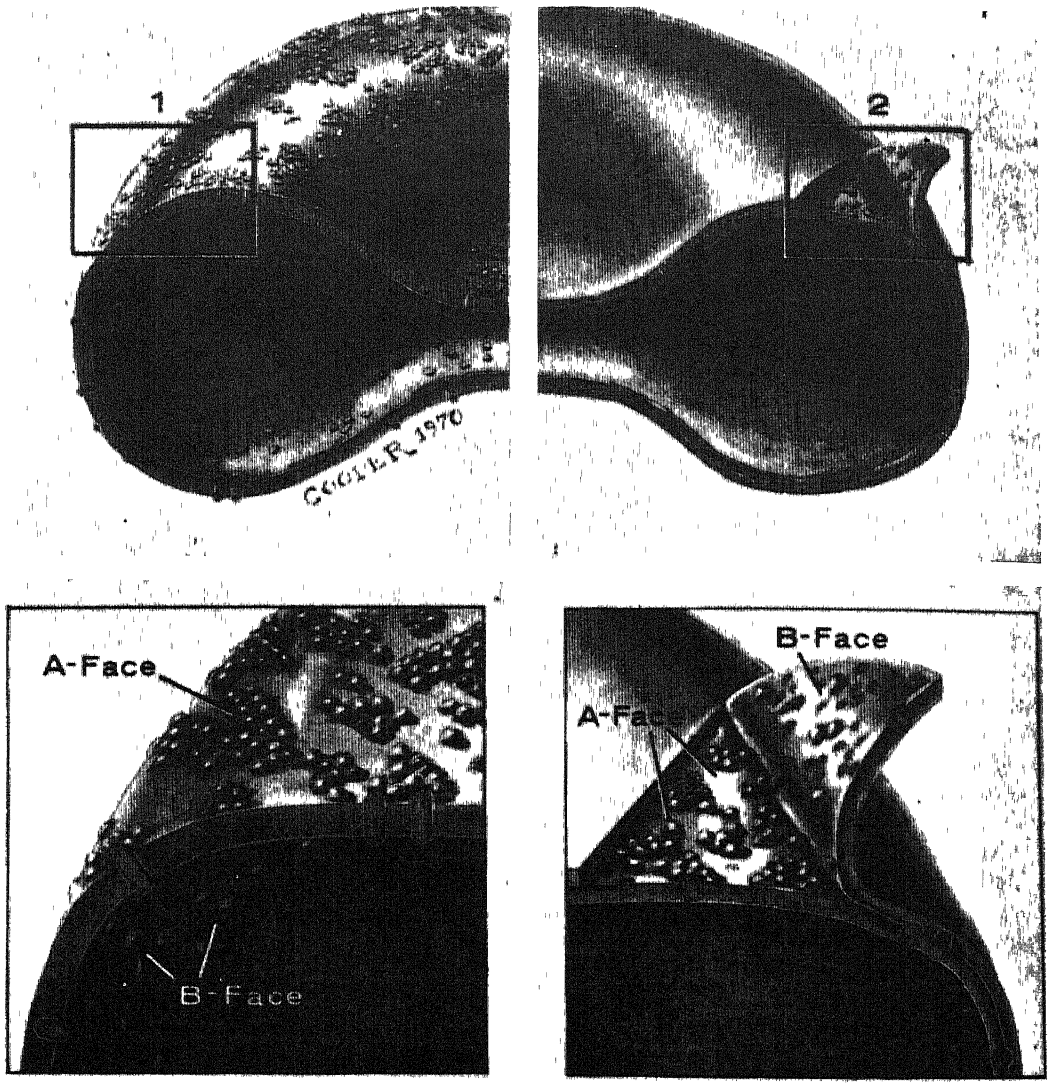


Fig 4 12 Artist's concept of the alternative interpretations of the location of the A-Surface and B-Surface of red cell membranes demonstrated with freeze-cleaving Interpretation 1 shows the membrane-associated particles on the true outer (*Face-A*) and inner (*Face-B*) surfaces of the membrane The two surfaces face in opposite directions Interpretation 2 shows the fracture entering the membrane and passing through its interior *Face-A* and *Face-B* point toward each other within the membrane (From Weinstein and McNutt 1970)

maintaining cellular integrity and the high internal concentration of potassium, and for prevention of the oxidation of heme is provided by these two pathways

a) The *Embden-Meyerhof pathway* carries out, through a series of phosphorylated intermediates, the anerobic conversion of glucose to lactate with the net production of potential energy in the form of two moles of adenosine triphosphate (ATP) for each mole of glucose The ATP's stored energy is essential for maintaining the normal cation concentration and balance of the red cells Reduced diphosphopyridine nucleotide (DPNH) is also produced in this pathway, it is essential for the function of methemoglobin reductase

b) The *phosphogluconate pathway* oxidizes 1 mole of glucose to carbon dioxide and water and produces 2 moles of a compound with a high reductive capacity, reduced triphosphopyridine nucleotide (TPNH) Red cell survival in the circulation is related to the action of

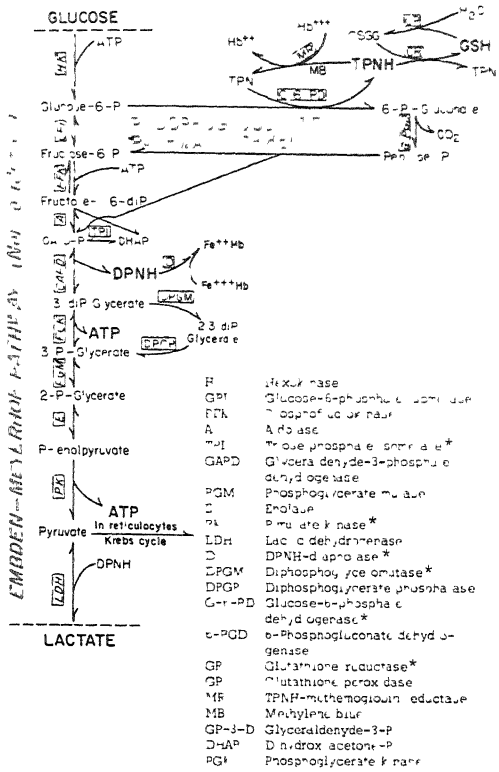


Fig 4-13 Erythrocyte metabolic pathways in adult erythrocytes Asterisks indicate recognized hereditary enzyme deficiencies In the glycolytic pathway deficiencies of pyruvate kinase triose phosphate isomerase and diphosphoglyceromutase are responsible for shortening cell survival In the oxidative pathway deficiencies of glucose-6-phosphate dehydrogenase glutathione reductase or GSH itself provide a liability to exogenous or endogenous oxidative stress and thus to acute hemolysis with oxidant drug exposure or if severe enough to chronic hemolysis without oxidant drug administration (From Simon et al 1966)

glutathione reductase upon thiols of the red cell membrane (see below)

The glycolytic pathway utilizes about 10 times as much glucose as the phosphogluconate shunt With aging of the adult red cell in the circulation, the activity of some of the enzymes of both pathways declines along with the lipids of the red cell envelope Certain deficiencies in the red cell enzymes are inherited and may produce severe hemolytic anemias, sensitivity of red cells to drugs, and short life span of the red cells It is beyond the scope of this chapter to go into details on the energy production and utilization in the red cell, but it seems to be reasonably well demonstrated that the energies developed are used for *1)* running the sodium and potassium pump to maintain the plasma red cell gradients *2)* keeping the hemoglobin in a functional state capable of reversible deoxygenation, and *3)* maintaining the normal equilibrium between oxidized and reduced glutathione participation in the maintenance of a lipid turnover and functions not yet described

Abnormalities in the phosphogluconate oxidative pathway are clinically important, even though only 10% of the glucose used is metabolized via this pathway Abnormalities in this shunt may result in acute or chronic hemolytic anemia The exact mechanism is not known but it is believed to be related to the function of this shunt in maintaining normal concentrations of thiol compounds in the red cell and its membrane One type of abnormality is a deficiency of glucose 6-phosphate dehydrogenase (G6PD), which may be present in as many as 100,000 000 persons The most severe of the hemolytic anemias are produced by drugs in individuals with G6PD deficiency In a few persons a chronic hemolysis may occur without prior drug sensitization

An example of a severe drug effect in G6PD deficiency is the condition that may occur in the treatment of malaria with Primaquine, where some individuals develop severe acute hemolytic anemia Negroes are more susceptible to the hemolytic effect of this agent than white people The susceptibility resides in the subject's red cells, since following the transfusion of tagged red cells from a sensitive subject into a nonsensitive subject, there occurs a striking reduction in their lifespan when Primaquine is then given to the nonsensitive recipient G6PD activity, reduced in the blood cells of Primaquine-sensitive individuals, is concerned with regeneration of TPNH and is required in the reduction of oxidized glutathione The glutathione level of sensitive cells is consistently lower than that of nonsensitive cells, and a rapid fall in the red cell glutathione level occurs in vivo when Primaquine is administered to sensitive individuals, but not when it is given to insensitive persons

Many other drugs may produce hemolysis in G6PD deficient individuals, rarely, even aspirin The observation that the eating of fava beans produces severe hemolytic anemia in G6PD deficient Caucasians but not in Negroes shows that at least two different biochemical defects exists

THE RED CELL LIFE SPAN AND ITS MEASUREMENT

The first reasonable estimate of the life span of the red cell was made by Ashby (1919) Her technique involved a transfusion of compatible but immunologically distinct red cells that could be identified by agglutination at regular intervals Techniques available now are the cohort labeling methods (glycine labeled with nitrogen-15, carbon-14, tritium or radioiron) and the random labels (immunological marker $DF^{32}P$, tritium, carbon-14, ^{51}Cr) The indirect methods of plasma radioiron clearance and red blood cell radioiron uptake, carbon monoxide excretion and bilirubin production are useful but more cumbersome (Berlin et al , 1959, Berk et al , 1970)

The *cohort method* depends upon the labeling of the red cell by biosynthetic incorporation of a precursor into hemoglobin in the marrow This method has the requirement that the precursor's specific activity must be decreased rapidly by degradation so that there is essentially a "flash" labeling of the cohort of cells One can then observe these cells as they emerge from the marrow, live out their life span and die

The *random method* involves the labeling of cells of the peripheral blood in vitro (irrespective of their age), followed by autotransfusion and then regular estimates of their survival

The *indirect methods* deduce the rate of synthesis of red cells by mathematical interpretation of iron kinetics or the rate of elimination of hemoglobin catabolites such as CO and bile pigments (Berk et al , 1970)

All of the methods for determining red cell survival depend on the assumption of a steady state equilibrium between entrance and exit of red cells from the blood The indirect methods are acutely sensitive to diversions from a steady-state equilibrium, and are applicable only when the total red cell mass is constant or where the rate of change of the total red cell volume is known

In Figure 4 14, examples of cohort and random labeling curves are shown The glycine cohort labeled curve discloses the emergence of labeled cells into the blood The specific activity rises to a maximum, and then falls very slowly for 3 to 4 months due to random loss, and then is terminated by a more rapid decrease in a falling sigmoid curve that represents the integral of the frequency distribution of the individual red cell life spans (Shemin and Rittenberg, 1946) The midpoint on the descending curve signifies the mean life span It cells are being produced with a shortened life span, the curve is shifted to the left If there is a random destruction of cells as they emerge into the peripheral blood, one sees an immediate diminution in the concentration of the cohort labeled cells, only a small fraction survives to attain the mean life span of the normal red cells The DFP curve is similar to an Ashby curve and represents the disappear ance of a population of cells that is random is respect to their age from birth The older cells commence disappearing promptly, and there is a linear decline as cells age with the curve crossing the axis at approximately the mean life span of the red cell in the blood If there is a shortened life span, the curve will be linear with a steeper slope With random destruction there will be a curvilinear disappearance The ^{51}Cr curve is curvilinear because there is a slow, steady elution of Cr from the labeled cell Although this complicates the interpretation, ^{51}Cr labeling with $Na_2{}^{51}CrO_4$ is the technique generally used clinically because of the simplicity of the procedures

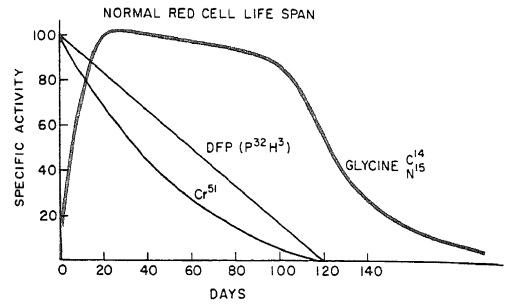


Fig 4 14 The specific activity of red cells in normal subjects after labeling with glycine DFP or ^{51}Cr (From Berk et al 1970)

VARIATIONS IN SIZE SHAPE, NUMBER AND STRUCTURE OF RED CELLS

Under physiological conditions, little change in shape of the red cell occurs although a few fragmented cells may be found in normal blood A slight change in volume, about 7 5% occurs due to osmotic changes incident to the respiratory cycle The red cells at birth and in early infancy are larger than in adult life This as well as the higher cell count is responsible

for the high packed cell volume (54%) of the infant's blood

In disease where the red cell is subject to many changes in size shape and structure examination of the blood and the identification of the various forms of abnormal cells is an important means employed in diagnosis of the different anemias Only the more outstanding abnormalities can be touched upon here An increase in the number of *reticulocytes* may be seen These young red cells resemble the ordinary cells in every way except that after supravital staining (alcohol solution of brilliant cresyl blue added to blood in the fresh state) a fine reticulum of basophilic material can be seen in the cytoplasm In normal blood, the reticulocytes are from 0 to 2% (av 0 8%) of the total red cell count They are increased after hemorrhage at high altitudes, by exercise, and in pernicious anemia following specific treatment A rise in the reticulocyte count indicates an increased activity of the blood forming tissue—the bone marrow—which as a result of a specific stimulus turns out a larger number of young cells

Shrinkage of the contents of the red cell (*crenation*) with wrinkling of its limiting membrane results from immersing the cells in a hypertonic solution In pernicious anemia, in particular, the presence of cells of unequal sizes (*anisocytosis*) and of deformed outline (*poikilocytosis*) is common The poikilocytes may assume the most bizarre forms, mulberry, target, flask, or hammer shapes may appear Erythrocytes with a globular form (*spherocytes*) or of an eliptoid or crescentic shape (*eliptocytes* and *sickle cells*) are found in certain types of anemia In sickle cell anemia, the red cells contain an abnormal type of hemoglobin (Pauling et al, 1949) In the deoxygenated form this is insoluble and forms crystalline "tactoids" within the cell membrane *Macrocytes* and *microcytes* are terms denoting red cells of larger and smaller size, respectively In certain conditions, e g, lead poisoning, fine dots of basophilic material appear throughout the cell, giving it a stippled appearance This abnormality is known as *punctate basophilia* In certain types of anemia, rings or twisted strands of basophilic material may be seen near the periphery of the cell These are derived from the nucleus and are known as Cabot's rings At other times small nuclear fragments Howell-Jolly bodies are present in the cytoplasm

Anemias

CLASSIFICATION ON BASIS OF RED CELL SIZE AND HEMOGLOBIN CONTENT

Mean Corpuscular Volume (MCV). This expresses the volume of the individual red cell in μm^3 (1 $mm^3 = 10^{12} \mu m^3$). The mean corpuscular volume in μm^3 is obtained by dividing the volume of packed cells expressed in ml/liter by the red cell count in millions/mm^3 (e g 450/5 - 90 μm^3)

Anemias may be classified according to the MCV as macrocytic (over 96) normocytic (86 to 96) and microcytic (under 80)

Mean corpuscular hemoglobin (MCH) is the hemoglobin per red cell expressed in $\mu\mu g$ (10^{-12} g) This may be obtained by dividing the hemoglobin per red cell expressed in $\mu\mu g$ (10^{-12} g) This may be obtained by dividing the The mean corpuscular hemoglobin may vary with the concentration of hemoglobin within the cell and with the volume of the cell

Mean corpuscular hemoglobin concentration (MCHC) is the hemoglobin in g/100 ml of blood divided by the packed cell volume of 100 ml expressed as a percentage e g

$$(15/45) \quad 100 = 33\%$$

Above 30% is normal (normochromic), below 30% is hypochromic

In Table 4 4, the normal values and standard errors of red cell dimensions are tabulated The hemoglobin is best determined by a cyanmethemoglobin method The volume of the

Table 4 4
Normal values for red blood cells

	Males	Females
Red cell count/mm^3	5 4 ± 0 8	4 8 ± 0 6
Hemoglobin (g/100 ml)	16 0 ± 2 0	14 0 ± 2 0
Volume packed red cells (%)	47 0 ± 7 0	42 0 ± 5 0
Mean corpuscular volume MCV (μ^3)	87 ± 5	87 ± 5
Mean corpuscular hemoglobin MCH ($\gamma\gamma$)	29 ± 2	29 ± 2
Mean corpuscular concentration MCHC (%)	34 ± 2	34 ± 2

From Wintrobe 1967

packed red blood cells is determined by centrifuging blood to a constant volume of the red cells Variations in these indices are very useful in studying physiological processes that influence erythropoiesis, and also in the diagnosis of disease In Figure 4 15 the frequency distribution of red cell volumes for normal and microcytic red cells is shown

The red cell concentration in the past was counted microscopically by direct vision in a hemocytometer Determination of the size distribution of red cells was a tedious microscopic procedure of measuring visually several hundred cells on stained smears Electronic gating methods now determine the number and volume of particles rapidly and quickly, and may yield a graph of the size distribution within 100 sec (Brecher et al , 1962) In addition, more sophisticated, automatic electronic methods determine hemoglobin and red count, and print out red cell constants

Examples of abnormal red cells are shown in Figure 4 16

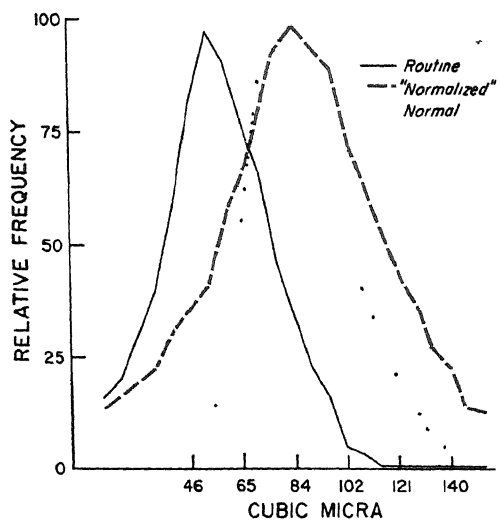


Fig 4 15 Normalization of microcytic curve for estimate of coefficient of variation (anisocytosis) Note microcytic curve has slightly narrower base than normal curve in routine run i e slightly smaller S D For estimate of coefficient of variation (C.V) however peak must be shifted by changing amplification to coincide with normal peak In the present case, C V is markedly greater than normal as can be seen from the much broader base of "normalized" curve (From Brecher et al 1962)

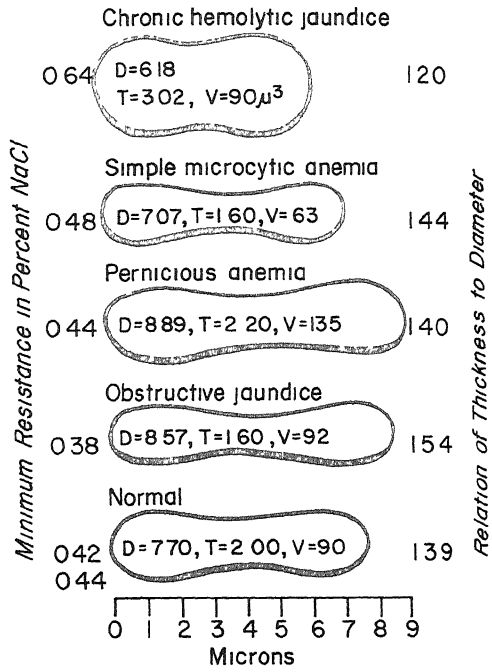


Fig 4 16 Red cell diameter thickness and volume in various clinical conditions compared with the normal *D* diameter *T* thickness *V* volume

HYPOCHROMIC MICROCYTIC ANEMIAS—IRON DEFICIENCY

In this group of anemias the essential defect is one of hemoglobin formation The hemoglobin percentage of the blood is reduced to a greater extent than the number of red cells The latter, indeed, may show only a slight reduction The MCH is considerably below the normal, which means that each red cell has received less than its normal quota of pigment The erythrocytes are also smaller than normal, so the MCV is also low The low MCH is in part the result of the smaller size of the red cell but also of a reduced concentration of pigment throughout the red cell's substance Some of the corpuscles are so pale that they either resemble "ghosts," or only the peripheral zone of the cell is colored (anisochromasia)

Iron deficiency is the cause of almost all the anemias belonging to this class The deficiency may result from increased iron requirements coupled with either a relative or absolute inadequacy of iron in the diet, or from defective absorption of the metal from the food Hypochromic anemia occurs most frequently in women of childbearing age in whom the dietary intake of iron is insufficient to meet their needs The anemia is aggravated by the losses of iron incident to menstruation or repeated pregnancies. In adult males the usual cause of this type of anemia is blood loss

Anemia of Infants. The fetus accumulates a

store of iron in the liver in the later months of gestation This serves as a reserve that is drawn upon for the manufacture of hemoglobin in infancy The high red cell concentration with which the infant comes into the world also contributes to the iron reserves In the normal infant the iron stores are sufficient for the manufacture of hemoglobin for the first 6 months or so Growth, however, makes heavy demands on the iron supplies, and after the first ½ year it is necessary to provide a diet that contains adequate amounts of iron in order to guard against the development of hypochromic anemia Cow's milk is very poor in both iron and copper If the iron stores are deficient at birth as in premature infants or as a result of maternal anemia, anemia may occur in the very young infant

The hypochromic anemias respond in a spectacular fashion to the administration of inorganic iron (e g , ferrous sulphate or carbonate, etc) The administration of copper is never necessary since this occurs in sufficient quantity in the diet and as an impurity in iron preparations

PERNICIOUS ANEMIA (ADDISON S ANEMIA) AND RELATED MACROCYTIC ANEMIAS

Pernicious anemia is due to a defect in the formation of the red blood cells Its chief features are

1) There is a great reduction in the number of red cells and consequently in the hemoglobin percentage The red blood cell counts in a very severe case may be less than 10% of the normal

2) The MCH is nevertheless elevated because the numbers of the red cells is reduced to a greater extent than is the hemoglobin percentage The MCHC is normal

3) The cells are large, the average diameter of the cells being increased to between 8 and 9 μm These large cells are called *macrocytes* The average volume of the individual red cells (Fig 4 16) is about 135 μm^3 (normal about 90 μm^3) *Normoblasts* and earlier forms may be present in the circulation , A characteristic large nucleated cell containing basophilic material in its cytoplasm and little or no hemoglobin is a feature of the bone marrow picture This cell, known as the *megaloblast*, differs chemically and morphologically from any cell of the normal erythrocyte series found either in bone marrow or blood

4) A great variation is found in the size of the cells—*anisocytosis*, the cells range from those smaller than normal to the large cells mentioned above *Poikilocytes* are markedly increased in frequency

5) There is an increase in iron and bilirubin of the plasma with increased excretion of pigment (urobilin, see below) The increase in plasma iron is due to its decreased utilization for the formation of red cells The increased plasma bilirubin is the result of increased red cell breakdown

6) Achlorhydria almost always exists

7) The tongue is sore and shiny, with atrophy of lingual papillae

8) There may be a combined degeneration of the spinal cord

9) There is an atrophy of the mucosa of the fundus and body of the stomach

In pernicious anemia the influx of cells into E_5 is only 5 to 15% of the K_{out} from E_4 indicating a very marked death in the production line In addition, the rate of replacement of E_5 by E_4 cells is drastically reduced from 2 91% to 0 3 to 1 67%/hr (Messner et al , 1969) An increased proportion of E_4 cells is found in G_2 (tetraploid DNA) and in cells between diploid and tetraploid that do not label with ^{3}HTdR in vitro and hence are shown to have stopped DNA synthesis (Wickramasinghe et al 1968) The in vivo and in vitro studies quite clearly demonstrate an abortion of cells in the production line This assertion is substantiated by the observation that this ineffective type of megaloblastic hemopoiesis declines for 2 to 2 5 days after vitamin B_{12} therapy during which it is replaced by normal erythropoiesis from the stem cell This interval sets perhaps a minimum time of 2 days for normal transit from the stem cell to the normal reticulocyte (Hillman et al , 1968) When erythropoiesis in guinea pigs is suppressed by induction of polycythemia there remain almost no reticulocytes in marrow or blood When the animals are then bled to induce new erythropoiesis, approximately 2 5 days are required for transit from the erythropoietin-sensitive stem cell to the reticulocyte (Griffiths et al , 1970)

The hematological aspects of this disease are cured by injections of vitamin B_{12} The progression of neurological complications is stopped but not reversed The pathogenesis will be discussed later

Red Cell Destruction, Normal and Abnormal

Under normal circumstances the red cells are subjected to trauma as they are deformed in passing through the capillary circulation For example, some of the splenic blood vessels may be only 3 μm in diameter compared to 7 μm for the red cell The capability of the red cell to undergo this deformation and return to the normal biconcave shape has been long appreciated Moreover, it can be deformed by relatively small forces The deformability and flexibility are exceptionally important in its physiological function Shear processes and deformation contribute to fragmentation of red cells Normal aging storage of red cells as in a blood bank, abnormalities in hemoglobin synthesis, sulfhydryl inhibitors, on inherited disorders of the membrane result in loss of intrinsic membrane deformability, or may alter the shape or the rigidity of cellular contents Within the spleen the pH may decrease and the cell membrane become stiffer, and this increases the probability of the cell's being removed from the circulation Alterations in the deformability of red cells appears to be a common precursor to destruction in all of the hemolytic anemias (LaCelle, 1970)

The destruction of red cells may take place intravascularly or extravascularly In the case of intravascular hemolysis, the hemoglobin is liberated and is bound by the haptoglobin The more usual type of destruction of red cells is extravascular and results from some alteration in the red cell membrane that is recognized by the reticuloendothelial system and causes the abnormal cells to be removed as they go through the spleen, lymph nodes, and bone marrow The most striking removal is by the spleen

The red cell defects that may lead to a short life because of cell removal may be due to an intracorpuscular abnormality that may be inherited or acquired Examples of the intracorpuscular abnormalities are hereditary spherocytosis, thalassemia, and hemoglobinopathies (about 40 have been described), sickel cell disease is a prime example Hereditary defects in carbohydrate metabolism also are important The extracorpuscular abnormalities are usually due to antibody coating of red cells, this is seen in transfusion reactions, erythroblastosis fetalis, or autoimmune hemolytic disease In addition, there is toxic nonimmunological hemolysis by phenylhydrazine, surface active compounds, heavy metals, infectious agents that produce hemolysins such as *Clostridum welchii*, malaria, other chronic infections, and malignancies There also may be combinations of intracorpuscular and extracorpuscular ab normalities

STRUCTURE OF THE SPLEEN RELATED TO REMOVAL OF AGING AND ABNORMAL RED CELLS

The first important fact about the spleen is that is constitutes a very large filtration bed In the normal adult about 2 liters of blood per minute pass through the spleen The structure of the organ is ideal for the detection and removal of abnormal or nondeformable cells (Fig 4 17) The splenic artery breaks up into trabecular arteries which continue to divide and penetrate into the parenchyma of the spleen Branches of a trabecular artery enter the white pulp of the spleen The dense white pulp consists of a network of loose reticular connective tissue whose interstices are packed with small lymphocytes and other cells This condensation of lymphocytes around the central arteriole is called the periarterial lymphatic sheath Some tiny branches of a central arteriole penetrate into and end in the lymphatic sheath However, most of the arteriolar branches penetrate through the sheath and terminate in the marginal zone or may continue into the red pulp In some areas there will be Malpighian corpuscles, these are produced by immunological stimulation and are shown to be sites of antibody production Some of the tiny arterioles that penetrate to the outside of the periarterial sheath of lym-

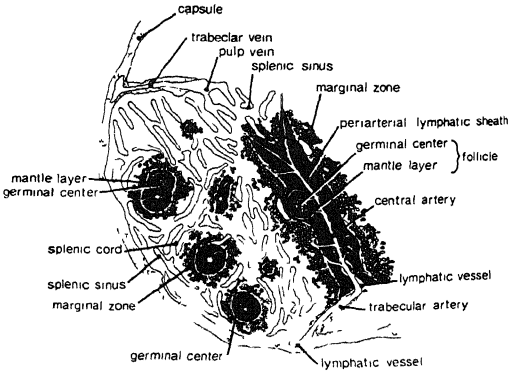


Fig 4 17 Plan of the structure of the spleen (From Weiss and Tavassoli 1970)

phocytes may end in splenic sinuses in an area called the marginal zone Other arterioles end blindly in the fine meshed reticular connective tissue of the marginal zone so that the cells of the blood migrate through the filtration network to the large number of venous sinuses in contact with the outer edge of the periarterial lymphatic sheath The marginal zone is continuous with the cords of the red pulp The reticular lacework of the marginal zone is jammed with macrophages, lymphocytes plasma cells and red cells The function of the marginal zone is filtration, as can be shown by the accumulation of particulate matter that has been injected into the splenic artery This area is filled with many phagocytic cells that will have the opportunity to recognize abnormalities and ingest these cells The central arteriole becomes smaller and finally enters and ends in the red pulp, rarely if ever connecting to venous sinuses The red pulp consists of vascular sinuses and cords The sinuses are thin walled venous vessels that anastomose into the venous system draining the red pulp

There are many hazards to cells traversing the spleen The splenic arterioles at their terminal ends are smaller than a red cell and possess a high endothelium between which the cell must squeeze Older cells are more fragile than young cells and may be broken in transit The less deformable cells, such as the spherocyte and sickle cells, may also be broken up Flow through the red pulp is irregular and there may be considerable stasis The stasis is associated with a decrease in oxygen content and pH, with a resulting shift in the oxygen dissociation curve that further lowers the oxygen tension All of these factors tend to cause the red cell to become more spherocytic and thus more susceptible to mechanical injury Antibody-coated cells are detected and removed by the phagocytes

Most red cells circulating through the spleen pass into the marginal zones and directly into the sinuses In doing this they pass through "windows" in the basement membrane and squeeze between the endothelial cells of the sinuses They may lose small bits of their cytoplasm in this transit and thus become smaller and more spheroidal and so less deformable on the next transit (Weiss, 1962, Weiss and Tavassoli, 1970).

It has been demonstrated that perfusion of the normal spleen by a mixture of normal and spherocytic red cells results in the removal of the spherocytes In addition, if one damages red cells by antibody or heat they are also trapped and removed in the spleen this can be determined by utilizing radio chromium labeled cells and external counting techniques Furthermore abnormal cells in transit through the spleen have inclusions removed from them Following splenectomy there is no detectable change in the life span of the red blood cell nevertheless, the number of abnormal red cells in the blood increases Consequently there is no doubt that the spleen is the major site for removal of aged and abnormal cells

CHEMICAL CHANGES RESULTING FROM INCREASED DEGRADATION OF HEMOGLOBIN

Hemoglobin degradation takes place within the entire reticuloendothelial system (A complete description is beyond the scope of this chapter) Globin is broken off the porphyrin ring is opened at the α-methene bridge, and iron is lost from the molecule but is very efficiently retained within the body for new hemoglobin synthesis When the porphyrin ring is opened at the α-methene bridge, 1 mole of carbon monoxide is formed for each mole of hemoglobin degradated (Nyman et al., 1959) The straight chain resulting from the cleavage of the porphyrin ring is reduced to bilirubin, a very insoluble material normally transported by an α-1 globulin When large amounts are produced the spill over is transported by albumin to the liver Within the liver, bilirubin is conjugated into glucuronide, a soluble form, that is excreted into the bile Normally a tiny amount leaks into the blood stream or perhaps is absorbed from the gastrointestinal tract The bilirubin glucuronide can be detected chemically because it reacts with Ehrlich's diazo reagent producing what is called a *direct reaction* showing the presence of conjugated bilirubin The unconjugated protein-bound material reacts only after treatment with alcohol to solubilize it and produce the color reaction This is called the *indirect reaction* Higher than normal indirect reaction indicates a hemolytic process Higher than normal direct reactions indicate an obstruction or malfunction of the liver and biliary tracts that permit leakage into the blood stream Within the gastrointestinal tract, the bilirubin is degraded

further by bacteria to a series of compounds known as urobilinogens Some of the material is reabsorbed into the blood stream but with the properly functioning liver they are rapidly cleared from the blood and little is found in the systemic circulation or the urine The presence of abnormally high amounts in the urine and stools is indicative of an increased hemoglobin catabolism

In normal man with red cell lifespan of 120 days, some 0 83% of his red cells per day are destroyed, representing about 6 g of hemoglobin Hemoglobin has a molecular weight of 66,000 and yields 4 molecules of urobilinogen Each gram of hemoglobin can produce 35 mg of urobilinogen The excretion of urobilinogen is theoxetically in the vicinity of 210 mg /day Substantial increases over this are indicative of a hemolytic process

RESPONSE TO HEMOLYTIC PROCESSES

Hemopoiesis is regulated primarily at the stem cell level When there is an increased peripheral destruction of cells from any hemolytic process there is a compensation within limits by increased red cell production that may prevent anemia The active red marrow may extend into the normally yellow fatty bone marrow For long periods of time the marrow can compensate for a reduction in life span of the red cell to one-sixth of the normal value This requires six times the normal production This is accomplished by an increased input of stem cells, possibly with a decrease in the generation time of the proliferating erythropoietic cells, and conceivably with an additional mitosis or two between the stem cell and the first nondividing red cell precursor A decrease in generation time of the red cell precursor at the expense of G_1 is conceivable It is doubtful whether S would be decreased In any case the reduction in time by itself could not accomplish a sixfold increase in production rate

When the red cell lifespan is shortened to less than one-sixth, it is certain that a hemolytic anemia will develop One sees a low hemoglobin and red count, the shortened red cell lifespan, increased reticulocyte count, possibly nucleated red cells in the peripheral blood, increased iron turnover rate and a markedly increased excretion of the products of hemoglobin destruction In addition, the extension of the marrow out into the long bones can be detected by radioisotopic techniques In severe sickle cell anemia and thalassemia the marrow may actually erode into the cortical bone producing typical x-ray pictures

EXAMPLES OF DISEASES WITH AN INTRACORPUSCULAR DEFECT

Hereditary Spherocytosis. This disease is inherited presumably as a single Mendelian dominant It is characterized by *1)* acholuric jaundice, *2)* fragile red cells spherocytic in shape, osmotically sensitive with intrinsic defective metabolism and structure, *3)* splenomegaly, and *4)* hemolysis with variable anemia The osmotically fragile red cells, being spheres, are not deformable and are easily removed in the transit through the spleen Thus, the lifespan of the red cell is randomly reduced, producing anemia if the marrow can not compensate for the shortened red cell life A large spleen develops, it is probably due to hyperplasia as a result of the extensive amount of red cells and hemoglobin that are being removed and degraded

The disease is always strikingly improved by splenectomy, because this operation removes a major hazard to the abnormal red cell The spleen itself, however, is not abnormal, since it can be shown by the Ashby technique of differential agglutination that normal cells transfused into a patient with hereditary spherocytosis both before and after splenectomy have a normal lifespan The defect then is related directly to an abnormality of the red cell

Thalassemia In Greece, a young married couple may bring to the Children's Clinic a child who is listless and pale with a prominent facies, and who has a hypochromic microcytic anemia with hyperplastic bone marrow There is elevated fetal hemoglobin and the other signs typical of the disease, *thalassemia major* This child will have a short and miserable life This disease is a national catastrophe in Greece and other parts of the Mediterranean littoral The only treatment is blood transfusion, but it will eventually cause trouble through "iron loading" from the destruction of the transfused red cells, and will lead to harmful deposits of iron (hemosiderosis) in liver, heart, brain, and other tissues.

This is an inherited disease Thalassemia major is the homozygous form, Thalassemia minor the heterozygous The latter gives relatively few symptoms, but the trait is detect-

able There is an intracorpuscular defect of the red cells, which have a very low MCV and MCHC There are abnormalities in hemoglobin synthesis and red cell formation and a large spleen Although there is no chemically defined abnormality in hemoglobin, the fetal type and A-2 variety of hemoglobin are present in abnormal quantities and at unusual ages Ineffective erythropoiesis is marked by red cell hyperplasia but inadequate red cell production The red cells show both an increased osmotic resistance and a marked shortening of lifespan upon transfusion into normal individuals or as measured in the patient In the homozygous form, the red cell lifespan is decreased so much that the marrow, even with extension into and erosion of the bones, is unable to keep up with cell destruction and a severe anemia occurs

Examples of extracorpuscular abnormalities that produce shortened red cell life span, such as mismatched transfusions and acquired hemolytic disease of the autoantibody type, are described in Chapters 6 and 7

Sickle Cell Anemia: A Molecular Hemolytic Disease. A negro child, normal at birth, developed somewhat slowly Intermittent periods of aching in the joints, severe chest and abdominal pain commenced in midchildhood An anemia developed The extremities were long and thin There was poor development of secondary sex characteristics This child has sickle cell disease, a chronic hereditary hemolytic disease due to the inheritance from each parent of a gene for hemoglobin S There is little, if any, normal hemoglobin, the fetal hemoglobin of red blood cells being replaced by hemoglobin S The pathology is molecular, It is based on a mutation with a substitution of valine for glutamic acid in the β chain of hemoglobin (Pauling et al, 1949, Ingram, 1956) The substitution involves only the β chain, and this explains why children with the homozygous form of the disease do not show the sickling phenomenon until adult hemoglobin S predominates over fetal hemoglobin (see chap 4) When the red cells are well oxygenated, the hemoglobin behaves physiologically But, when these cells are deprived of oxygen they assume sickle and other bizarre crescentic shapes When hemoglobin S is deoxygenated its solubility falls strikingly (Harris, 1959) This decrease in solubility results in a viscous semisolid cell that loses almost all of its deformability The rigid cells become trapped in the smaller blood vessels, where stasis further deoxygenation and reduced pH accentuate the situation Plugs and masses of erythrocytes occlude vessels and thrombosis and infarction follow A vicious cycle thus leads to more sickling The clinical results are severe painful crises resulting in multiple infarcts of bone, spleen, liver and to chronic leg ulcers (The repeated infarcts of the spleen may lead to an autosplenectomy) A chronic compensated, hemolytic anemia develops When the anemia becomes severe, the only therapy is blood transfusion The latter ultimately leads to difficulties through the increased iron intake as described in thalassemia Splenectomy does not improve the disease

A hazard to the individual with sickle cell disease is the possibility that a mistaken diagnosis of the painful crises may lead to emergency and unnecessary surgery

Pathogenesis of Pernicious and Megaloblastic Anemias

Whipple and his associates received the Nobel Prize in 1935 for having demonstrated that liver is the most effective food for the treatment of anemia in dogs Minot and Murphy (1926) gave lightly cooked liver (up to 1 lb/day) to patients with pernicious anemia with considerable improvement Shortly thereafter, it was found that meat incubated with normal gastric juice and then fed to a subject with pernicious anemia produced a striking effect (Castle et al, 1930) My introduction to hematology as a medical student and house officer was concerned with incubating hamburger with normal gastric juice and then feeding it to patients with pernicious anemia by stomach tube Today one gives a simple injection of vitamin B_{12} There has been progress for patient and physician

INTRINSIC AND EXTRINSIC FACTORS VITAMIN B_{12}

In pernicious anemia, there is an abnormality in gastric mucosa with a reduced volume of gastric secretions (achylia) and absence of hydrochloric acid in these secretions Also there is a lack of an *intrinsic factor* which normally acts on *extrinsic factor* that is present in meat and other foods. The extrinsic factor is now known to be vitamin B_{12}, a red-

dish compound possessing a cobalt-containing tetrapyrrolic ring that is synthesized by various bacteria It is available to man in meat from animals whose intestinal bacteria synthesized the material or consumed it in their diet The efficient absorption of dietary vitamin B_{12} requires an interaction with a glycoprotein of molecular weight 70,000 secreted by the fundus of the stomach, the intrinsic factor This factor binds strongly to vitamin B_{12} The complex then attaches to specific receptors that are on the intestinal wall and that facilitate the transfer of the vitamin B_{12} to the circulation, primarily through the upper part of the ileum In a normal individual with adequate amounts of intrinsic factor, approximately 70% of the ingested vitamin B_{12} is absorbed, whereas, in individuals such as those with pernicious anemia who have an absence of the intrinsic factor, less than 2% is absorbed The vitamin B_{12} that is absorbed through the gut is carried through the bloodstream by a specific protein carrier The serum concentration of vitamin B_{12} in normal individuals is 200 to 900 $\mu\mu$g /ml The total body vitamin B_{12} content in the normal adult is about 5,000 μg , about 1700 μg is stored in the liver Symptoms characteristic of pernicious anemia described earlier do not begin to appear until the total body B_{12} content drops to less than 500 μg The average daily requirement for vitamin B_{12} in the normal adult is approximately 2 5 μg During growth and severe demands on hemopoiesis the requirements may be increased It is beyond the scope of this section to describe the entire biochemical mechanism by which vitamin B_{12} exerts its effect on DNA synthesis and cell replication so as to eliminate the marked abnormalities in morphology and the inability of the cells to progress through the cell cycle that results in the marked ineffective erythropoiesis described earlier Since large amounts of thymidine will give a partial remission in pernicious anemia, it is believed that the defect may be focused on the synthesis of the DNA precursor, thymidylic acid.

The specific diagnosis of pernicious anemia requires the demonstration of the absence of intrinsic factor This is accomplished by using radioactive cobalt labeled vitamin B_{12} When the material is given without intrinsic factor, if there is no absorption into the body it is presumptive evidence of an absence of intrinsic factor Next, the radioactive vitamin B_{12} is incubated with a preparation of intrinsic factor and then given by mouth to the individual and the absorption measured again If now there is absorption, it is definite evidence of a deficiency of intrinsic factor

The technique of determining vitamin B_{12} absorption described by Schilling (1953) required the oral administration of radioactive vitamin B_{12} followed at an interval by a large quantity of nonradioactive vitamin B_{12} given intravenously to flush the radioactive material into the urine, where one estimated absorption on the basis of urinary excretion The alternative was to collect feces for a few days and measure the amount excreted The whole body counter makes it much simpler Cobalt-60 vitamin B_{12} is given by mouth followed by water to wash out the stomach Next cobalt-57 vitamin B_{12} is incubated with intrinsic factor and given by mouth The total body cobalt 60 and 57 are determined A few days later the radionuclide content of each in the body is determined If cobalt-60 has been absorbed in substantial amounts, the patient has his own intrinsic factor If only cobalt-57 was absorbed the patient has an absence of intrinsic factor If neither was absorbed significantly there is a defect in the small bowel that prevents absorption (Reizenstein et al, 1961)

FOLATES

The folates are also essential in DNA synthesis and normal cell proliferation In the absence of folates or inability to absorb them, macrocytic anemias may develop also The folates are synthesized by plants and are found in many vegetables and also in liver Most of the natural folic acid is in conjugated form and must be split by intestinal conjugases for absorption to proceed A normal diet contains daily from 1 to 2 mg of folic acid It is estimated that the normal daily requirement is about 50 μg During marked hemolysis, pregnancy, etc , the requirements are much higher Deficiencies develop in severe alcoholics

Folates are reduced by specific enzymes to tetrahydrofolic acid which accepts methyl groups and acts as a methyl transfer agent and coenzyme in the synthesis of some purines and pyrimidines required by DNA

The abnormalities in DNA synthesis induced by deficiency of vitamin B_{12} or folate affect all of the cell renewal systems The appearance is abnormal as shown by the mega-

loblast and the hypersegmented neutrophils in the blood Abnormalities in cell division in the other cell renewal systems account for the changes in the mouth stomach, vagina and skin The neurological damage is related specifically to vitamin B_{12} but how the vitamin B_{12} deficiency results in injury to neurons is not known

IRON METABOLISM

Iron is essential for the synthesis of hemoglobin and iron containing porphyrin enzymes The major abnormal loss of iron comes through hemorrhage If an individual must receive transfusions because of ineffective hemopoiesis due to suppression of the marrow or inadequate production of red cells, the total body iron content continues to increase in proportion to the number of transfusions given The only method normally available for regulating the total body iron content is through the regulation of intestinal absorption, there is no mechanism for regulating total body iron content by increasing its excretion

Most of the iron is absorbed through the upper jejunum, although small amounts can be absorbed anywhere from the stomach to the colon A considerable amount of the iron in food is released by peptic digestion in the stomach, and in the presence of the high acid content becomes ionizable inorganic iron (Jacobs and Greenman, 1969, Hoglund 1970) As the gastric contents flow into the jejunum and are neutralized, the ionized iron binds with other substances present in the lumen or is precipitated Iron complexes with phosphate, and other substances that are nonabsorbable may be produced. It is believed that some of the inorganic iron as ions is bound to gastric mucopolysaccharide which appears to act as a carrier enabling iron to remain in solution at the neutral pH in the small intestinal lumen (Jacobs and Miles, 1969) Heme iron of food is released from hemoglobin by digestion in the stomach, and apparently enters intestinal epithelial cells at which point it is degraded and the iron released to form intraepithelial complexes

The rate of iron absorption is regulated by the needs of the body A mucosal "block" theory has proposed that a regulatory mechanism for iron absorption resides entirely within the mucosal cells It was believed that the iron was bound by apoferritin into a highly insoluble form ferritin which was released according to the need for iron within the body (Granick 1946) The notion of regulation at the epithelial cell level in the small intestine has been questioned (Brown, et al, 1958), and has been replaced by the concept that control resides primarily at the level of the mucosal cell (Conrad and Crosby 1963)

It appears clear that individuals who are iron deficient increase their absorption of iron, whereas those with iron overload absorb relatively little Iron that has entered the epithelial cell at the luminal pole has two possible fates It may be transported to the base of the epithelial cell, across the basement membrane and into the plasma where it is bound by plasma transferrin The other route appears to be conversion to ferritin and its loss into the gastrointestinal tract as the mucosal cells are sloughed off of the tips of the villi [3] It is now believed that the prime regulator of the amount of iron absorbed into the body via the rapid transport method to the plasma is the capability of the cell to synthesize ferritin, sequester the gut iron, and then reject it into the intestinal lumen as the cells slough off (Conrad et al, 1964) It is not exactly established what switches on and off the synthesis of apoferritin to sequester the iron Certainly if there is a low mucosal apoferritin content the probability of iron's making the transit across the epithelial cell to be picked up by transferrin in plasma is much greater

Iron is transported in the blood by a single plasma protein, transferrin Approximately 30 to 40 mg of iron each day is required in a normal individual to synthesize hemoglobin This is 6 to 10 times the total plasma iron As described in Chapter 1, transferrin can bind two atoms of ferric iron per molecule The normal plasma iron concentration is between 60 and 150 μg/100 ml, which is about one-third of the plasma total iron-binding capacity This means approximately one-third of the circulating transferrin is saturated with iron

It is probable that transferrin molecules with

[3] The gastrointestinal tract is a cell renewal system in which the cells are being continually formed and desquamated The transit time from formation in the crypts of Lieberkuhn to the tip of the villus and being sluffed off is between 1 and 3 days depending on the species and the area of the gut

iron attach to developing red cells, allowing the iron to be transferred into the erythroblast with release of the transferrin to return and deliver more iron (Jandl and Katz, 1963).

The rates of iron transport, turnover, and incorporation into red cells are useful observations as indirect measures of red cell production. They are clinically and experimentally useful. The student is referred to hematology and nuclear medicine tests for details of techniques and data interpretation of iron kinetics (Wintrobe, 1967; Wagner, 1968).

As mentioned earlier, the commonest disease of iron metabolism is a deficiency induced by chronic hemorrhage in the male, repeated pregnancies and menstruation in the female, and nutritional deficiency of iron. There is an exceptionally rare congenital absence of transferrin. In chronic infection there is an ample amount of iron in the reticuloendothelial cells of the bone marrow, liver and spleen; there is a low serum iron and a low total iron-binding capacity along with a normocytic normochromic anemia. The exact pathogenesis is not clear, but removal of the infection usually completely corrects the disorder.

There is also a rare disorder of iron loading called hemochromatosis. The pathogenesis of this disorder is not clear, but it is characterized by extensive deposits of iron in the liver, skin, pancreas; any organ may be involved. Diabetes mellitus may develop as the result of injury to the islets of Langerhans. When the disease is fully developed, iron absorption is normal but is believed to be high in the early stages. Presumably it is due to a defect in the normal mechanism that controls iron absorption. The treatment of this disease is simply bleeding in order to remove the excess iron. This strikingly decreases the mortality of the disease by mobilizing the harmful deposits of iron in the tissues for erythropoiesis. The synthesis of heme and globin are equally important to erythropoiesis. The student is referred to standard texts on biochemistry for consideration of this aspect of erythropoiesis.

4 / HEMOGLOBIN[1]

Hemoglobin the red matter of erythrocytes makes up about 95% of the dry weight of the red cell Hemoglobin functions primarily as a carrier of oxygen but it is also concerned with the transport of carbon dioxide and with the acid-base balance of the body Hemoglobin takes up a large volume of oxygen in the lungs and carries it to the tissues Water at body temperature and an oxygen pressure of 100 mm Hg absorbs 0 3 ml of oxygen per 100 ml whereas 100 ml of normal blood at the same temperature and pressure takes up about 20 ml of oxygen The difference is due to the reversible oxygen-binding capacity of hemoglobin (Harper, 1963, Harris, 1970)

The blood's capacity to absorb oxygen is called the *oxygen capacity* and it is proportional to the hemoglobin concentration The oxygen capacity of 1 g of completely oxygenated hemoglobin is 1 39 ml therefore, the oxygen capacity of 100 ml of normal blood (with 15 g of hemoglobin) is 20 8 ml (15 × 1 39) (Gambino, 1967)

Blood from a normal male contains about 15 g (14 to 16 g) of hemoglobin per 100 ml , all of which is within the red cells Hemoglobin is 0 347% iron by weight, thus, there is about 50 mg of iron/100 ml of blood and 2 5 g in the total blood volume of 5000 ml (Gambino 1967)

THE ESTIMATION OF HEMOGLOBIN IN BLOOD

Hemoglobin concentration in blood can be measured by colorimetric, spectrophotometric, gasometric, chemical, and physical techniques, including x-ray fluorescence and atomic absorption spectroscopy (Drabkin, 1950, Gambino, 1967, MacFate, 1964) Photoelectric colorimetry determines hemoglobin concentration by measuring color intensity in terms of the light-absorbing power of the solution at a specific region of the visible spectrum The estimation of cyanmethemoglobin gives a rapid precise measurement of all hemoglobin derivatives except sulfhemoglobin

Gasometric methods measure hemoglobin through its ability to combine stoichiometrically with oxygen In principle, a measured volume of blood is equilibrated with air under standard conditions of temperature and pressure and is then analyzed in the Van Slyke blood-gas apparatus for its content of oxygen The results of gasometric methods represent functional hemoglobin in the blood i e that portion capable of carrying oxygen therefore, the nonfunctional pigments like methemoglobin and carboxyhemoglobin are not measured

Structure and Metabolism

Hemoglobin is a conjugated protein with a molecular weight of 64,458 (Gambino 1967) It consists of an iron-containing pigment (metalloporphyrin) combined with a protein, *globin* This pigment belongs to a group of compounds called *porphyrins* which combine with metals *Heme* is the name given to the compound of iron and the particular porphyrin found in hemoglobin The iron-porphyrin-globin complex (hemoglobin) forms a loose complex with oxygen when the iron is in the *ferrous* (Fe^{++}) state

SYNTHESIS OF PORPHYRINS AND HEME

The basic framework upon which all porphyrins are built consists of four pyrrole rings linked together in a larger ring-like structure called *porphin* By substitution of the 8 hydrogen atoms about the edge of the porphin structure with various side chains—acetyl, methyl, vinyl, propionyl—the various porphyrins in nature are produced The positions of some substituents for a few related structures are summarized in Table 4 5 The particular compound found in heme is known as *protoporphyrin IX* (MacFate, 1964)

The complex molecule is entirely synthesized by a series of condensations, beginning with glycine and succinate. From these simple

[1] See also Regeneration of Blood, Chapter 2

Table 4 5
Porphyrin structure

Compound	Side Chains
Uroporphyrin I	Acetyl in 1, 3, 5, 7, propionyl in 2, 4, 6, 8
Uroporphyrin III	Acetyl in 1, 3, 5, 8, propionyl in 2, 4, 6, 7
Coproporphyrin I	Methyl in 1, 3, 5, 7, propionyl in 2, 4, 6, 8
Coproporphyrin III	Methyl in 1, 3, 5, 8, propionyl in 2, 4, 6, 7
Protoporphyrin IX	Methyl in 1, 3, 5, 8, propionyl in 6, 7, vinyl in 2, 4

structures Δ-aminolevulenic acid (ALA) is formed, after which 2 molecules of ALA condense to form porphobilinogen The precise mechanisms of the subsequent reactions whereby 4 molecules of porphobilinogen condense to form protoporphyrin IX are not yet completely understood The protoporphyrin IX is combined with iron and globin to form a complete hemoglobin molecule *Iron* enters the red cell precursors through the cell membrane and is transferred to the site of heme synthesis where it combines with the protoporphyrin under the influence of *heme synthetase*

The breakdown of heme, when it is released from the red cell in the course of its normal destruction, involves the oxidative splitting of the porphyrin ring to yield bilirubin-urobilinogen products The conversion of hemoglobin to bilirubin probably takes place in the entire reticuloendothelial system, but the extent of participation by various organs (spleen, liver, bone marrow, etc) is unknown It is not certain whether *1*) the iron-porphyrin complex is split from the globin to form hematin (a hydroxide of the trivalent iron derivative of (heme), at which time the porphyrin ring opens and the iron is released to form the straight-chain bilirubin compounds, or 2) oxidation of hemoglobin (as the initial step) opens the porphyrin ring to yield choleglobin, a green, bile pigment that is an iron-globin complex, which upon subsequent loss of the iron and globin forms bilirubin (Harper, 1963, Harris, 1970)

By either route, the opening of the porphyrin ring occurs at one of the α-methene bridges, and 1 mole of carbon monoxide is formed Because endogenous production of carboxyhemoglobin occurs via hemoglobin degradation, the measurement of carbon monoxide production has been used as an indication of red cell survival (Harris, 1970)

After bilirubin, the principal bile pigment, is formed it is excreted into the gut A small but unknown portion of the bilirubin may be excreted either as mono- or dipyrroles, but it has been technically difficult to detect or quantitate these In the gut, bacterial action progressively reduces the bilirubin to members of the urobilin-urobilinogen groups Measurements of bilirubin production and fecal urobilinogen excretion have also been used to estimate hemoglobin degradation The possible degradation of cellular hemoglobin through a nonbilirubin pathway has been described, the hemoglobin may be degraded to iron and a porphyrin compound (Veech et al , 1967)

THE STRUCTURE OF GLOBIN

The hemoglobin molecule consists of four subunits, and each subunit consists of a polypeptide chain and an associated heme molecule The tortuously coiled polypeptide chains fit together in the complete molecule in such a fashion that the heme portions are more or less at the corners of a tetrahedron, near the surface of the molecule, where there can be a prompt reaction with oxygen Attachment of the heme portion greatly modifies the properties of the globin molecule The conjugated protein is much more resistant to denaturing agents and to digestive enzymes than is the free protein

Differences in various human hemoglobins (see below) are due to changes in the amino acid sequence of the polypeptide chains, or to different combinations of polypeptide chains A functional hemoglobin must have two pairs of different polypeptide chains In the few instances when a hemoglobin has only a single chain, the molecule does not function normally (Huisman and Schroeder, 1970)

Normal adult hemoglobin (A) has two so-called α-chains and two so-called β-chains and may be represented symbolically as $\alpha_2\beta_2$ The differences between the α-chains and β-chains lie in their primary structure each α-chain contains 141 amino acid residues, whereas, each β-chain contains 146 amino acid residues Fetal hemoglobin (F) (that normally present during intrauterine life) has one pair of α-chains identical to that in adult hemoglobin and a second pair of γ-chains, hence, it is

shown as $\alpha_2\gamma_2$ A minor fraction (2%) of hemoglobin normally present in adult human blood termed hemoglobin A_2 is made of two α-chains and two δ (delta)-chains, so it is shown as $\alpha_2\delta_2$ Hemoglobin H, with the structure β_4 and hemoglobin Barts, with the structure γ_4, are two rare abnormal hemoglobins (Huisman and Schroeder, 1970 Ingram 1963, Jonxis and Delafresnaye 1959)

The Combination of Hemoglobin with Oxygen

Hemoglobin combines with oxygen by virtue of the iron it contains Each of the four iron atoms in the heme groups reacts directly with molecular oxygen For this reaction to occur the iron must be in the divalent (ferrous Fe^{++}) state Deoxygenated hemoglobin has been shown to distort a magnetic field presumably because of unshared electrons on the four iron atoms On combination with oxygen, the unshared electrons participate in formation of the iron-oxygen bond and the molecule loses its magnetic moment When the iron atom is oxidized to the trivalent (ferric, Fe^{+++}) state, as in methemoglobin, it does not react with oxygen, although it will bind certain ligands such as cyanide Under normal physiological conditions, in vivo, the methemoglobin reductase system of the red cell maintains the iron of hemoglobin in the divalent state

In the hemoglobin molecule the iron is in a hexacovalent state, with four of the bonds linking it with the planar porphyrin molecule (Fig 4 18) The resulting heme molecules (4 in number) are attached to the surface of the globin molecule by another of the covalent links of the iron atom The sixth covalent bond which protrudes from the plane of the heme is the point of attachment of molecular oxygen It is evident that globin is not merely an inert carrier of the heme groups, on the contrary it markedly influences the reactivity of the iron of heme with oxygen (Perutz et al, 1968, Perutz 1969)

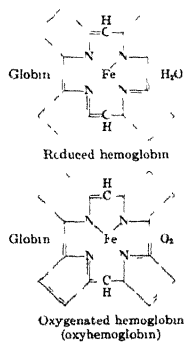


Fig 4 18

HEME-HEME INTERACTION

Since each hemoglobin molecule contains four iron atoms it can combine with four molecules of oxygen The oxygenation of one of the four hemes enhances the oxygen affinity of the remaining hemes This heme-heme interaction i e the effect of the oxygenation of one heme upon the ease of oxygenation of the other hemes results in a sigmoid curve that describes the percentage of oxyhemoglobin at a given oxygen tension This plot is called the oxyhemoglobin dissociation curve When there is no interaction such as in the case of the monomeric myoglobin, a similar plot will yield a curve in the form of a rectangular hyperbola (Antonini, 1965, Antonini and Brunori, 1970)

BOHR EFFECT

The reversible combination of hemoglobin and oxygen is dependent upon temperature, and on the ionic strength (electrolyte composition) of the reaction medium as well as on its P_{CO_2} and pH (Antonini, 1965, Antonini and Brunori, 1970) The interrelated P_{CO_2}-pH influence on the oxygen-hemoglobin equilibrium is known as the Bohr effect—i e , the higher the P_{CO_2} and thus the greater the hydrogen ion concentration (and the lower the pH), the lower is the affinity of the hemoglobin molecule for oxygen Figure 4 19 shows the effects of P_{CO_2} and pH on the red cell oxyhemoglobin dissociation curve The effect of one variable is shown when the others are constant The importance of the Bohr effect lies not only in its physiological role as a regulatory mechanism in the oxygen transport and gas exchange, but also in its use as a model for studying the effects of pH on the functional properties of proteins in general When hemoglobin is combined with oxygen, an imidazole group of the amino acid, histidine, in the α- and β-chains is influenced by the H^+ concentration, an increased H^+ concentration tends to promote dissociation of oxygen from hemoglobin

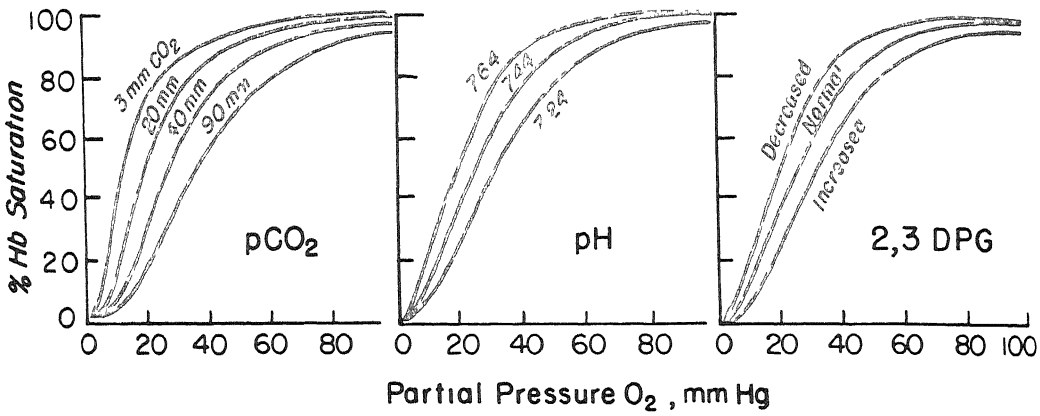


Fig 4 19 The influence of P_{CO_2} pH and 2 3 DPG on oxygen dissociation curves

THE BINDING OF ORGANIC PHOSPHATES TO HEMOGLOBIN

The oxyhemoglobin dissociation curve of extracorpuscular hemoglobin is different from that of cellular hemoglobin This difference may be explained by the existence of certain compounds with which hemoglobin is associated intracellularly

Organic phosphates bind to hemoglobin, and this binding is accompanied by a decrease in the hemoglobin molecule's affinity for oxygen Decreased hemoglobin-oxygen affinity is caused by many organic phosphates and by some inorganic phosphates, but the most dramatic results occur principally with the binding of 2,3-diphospho-D-glyceric acid (DPG), adenosine triphosphate (ATP), and inositol hexaphosphate (IHP) (Benesch and Benesch, 1969, Chanutin and Curnish, 1967) Because DPG and to a lesser extent ATP are present in significant amounts in human erythrocytes, their binding to hemoglobin is considered to be important to the function of the molecule in vivo Figure 4 19 shows also the effect of 2,3 DPG on the red cell oxyhemoglobin dissociation curve

The oxyhemoglobin dissociation curve of intact red cells is a resultant of oxygen-hemoglobin interaction influenced by a variety of intracellular substances -ionic concentration of hydrogen, sodium, potassium, and chloride, red cell levels of P_{CO_2}, 2,3 DPG, ATP, and glutathione, and the amount of carboxyhemoglobin and methemoglobin in the red cells (Antonini, 1965; Antonini and Brunori, 1970) Increased affinity of red cell hemoglobin for oxygen is associated with increased levels of red cell carboxyhemoglobin and methemoglobin, decreased concentration of hydrogen ion (increased pH), and decreased levels of red cell P_{CO_2}, 2,3 DPG, ATP, and glutathione Decreased affinity of the red cell hemoglobin for oxygen is associated with elevated red cell levels of 2,3 DPG, ATP, glutathione, and hydrogen ion (decreased pH) At a given P_{O_2} tension, decreased affinity for oxygen by red cells increases (facilitates) oxygen delivery to tissue, while increased affinity for oxygen decreases (reduces) oxygen delivery to tissue

Although the presence of DPG markedly reduces the oxygen affinity of hemoglobin, many variables determine the magnitude and nature of the effect The binding of DPG to hemoglobin is sensitive to a wide variety of factors There is a lower binding of DPG to hemoglobin F than to hemoglobin A DPG apparently controls the ease or difficulty with which hemoglobin releases oxygen to the tissues, thereby affecting the "position" of the oxyhemoglobin dissociation curve For example, in patients in whom defects in red cell glycolysis alter the DPG concentration in the red cell, there is abnormal oxygen affinity In a patient with hexokinase deficiency, the level of DPG is about two-thirds normal and the oxygen affinity is greater than normal, by contrast, in a patient with a deficiency of pyruvate kinase, DPG is more than double the normal level and the oxygen affinity is decreased

Tissue hypoxia plays an important part in regulating the red cell DPG level. Hypoxic hypoxia (pulmonary), stagnant hypoxia (cardiovascular), and anemic hypoxia (red cell mass deficit) all tend to raise red cell organic phosphates in patients with normal or elevated blood pH, a change that affects the oxyhemoglobin dissociation curve so as to favor unloading of oxygen to the tissues. Ascent of normal subjects to a height of 4500 m results in a reduction in oxygen affinity with a concomitant rise in red cell DPG level. The maximum effect is reached within 48 hr and values return to normal within 48 hr of descent to sea level. The mechanisms responsible for the changes in red cell DPG and ATP in vivo during hypoxic states are clearly linked to the red cell pH.

SIGNIFICANCE OF GLOBIN VARIANTS

Variations in the globin portion of the hemoglobin molecule influence the Bohr effect, the heme-heme interaction, and the hemoglobin molecule's affinity for oxygen in different ways. Globin variants in the α- and β-chains produce a large number of abnormal hemoglobins (hemoglobinopathies). The structural analyses of these abnormal hemoglobins yield important information about structural-functional relations in the hemoglobin molecule (Weatherall and Clegg, 1969). Some of the hemoglobin variants (Chesapeake, J Capetown, Yakima, Kempsey, Rainier, Yps., Hiroshima, and St. Mary's) are associated with increased oxygen affinity and usually with increased red cell volumes (familial polycythemia). Other hemoglobin variants (Seattle, Kansas, and the hemoglobins SA, SS, E, SO, and SC) are associated with decreased oxygen affinity and usually with reduced red cell volumes. Figure 4-20 schematically demonstrates the relation between hemoglobin mass and the oxyhemoglobin dissociation curve P50 value (partial pressure of oxygen in millimeters of mercury at which 50% of the hemoglobin is oxygenated) in patients with a variety of hemoglobinopathies.

Oxyhemoglobin dissociation function is an important factor in the regulation of oxygen transport to tissue. An increase in red cell 2,3 DPG in circulating red cells may tend to compensate for an impairment in oxygen supply to

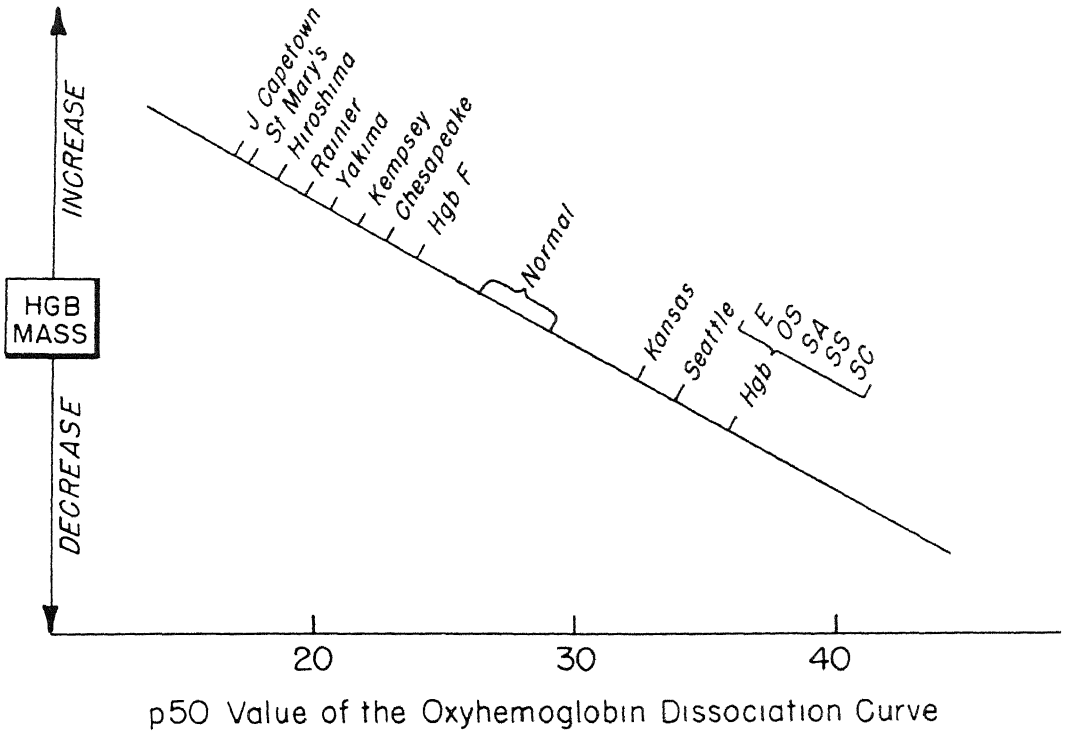


Fig 4-20 Relationship between hemoglobin mass and P50 of oxyhemoglobin dissociation curve for abnormal hemoglobins

tissue The adaptive mechanisms possessed by the body to assure an adequate oxygen supply in patients with either decreased red cell mass, cardiac disease, or pulmonary insufficiency include not only increased releases of oxygen from hemoglobin but augmentation of cardiac output, greater flow to more important organs, and usually an increased rate of red cell production

Derivatives of Hemoglobin

The terms, *reduced hemoglobin, ferrohemoglobin, or deoxyhemoglobin,* are used as synonyms for the molecule that has given up its oxygen

Oxygenated hemoglobin or *oxyhemoglobin* is applied to the molecule that has oxygen molecules attached

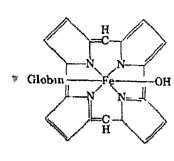
Methemoglobin, or *ferrihemoglobin* results when, by the action of oxidizing agents, the ferrous iron in the heme is changed to ferric iron (Fig 4 21) Normal blood contains about 0 1% methemoglobin, but in cases of poisoning by drugs such as nitrites, chlorates, sulfanilamides, etc , the percentage is increased So-called "toxic" cyanosis occurs when the more darkly colored methemoglobin amounts to about 3 g /100 ml of blood Methylene blue, injected into the blood stream, causes the formation of methemoglobin

A *hemochromogen* of hemoglobin results when the heme containing the ferrous iron is combined with a denatured globin

Cathemoglobin is a compound of heme containing ferric iron with denatured globin

Each of the various heme pigments and derivatives of hemoglobin has a characteristic absorption spectrum, thus, they may be readily identified and measured by appropriate spectrophotometric techniques The major absorption bands of a few compounds are listed in Table 4 6

Carbon Dioxide. Hemoglobin is of considerable importance in the carriage of carbon dioxide By virtue of the Bohr effect (see above)



Methemoglobin

Fig 4 21

Table 4 6
Wavelengths at the points of maximum intensity of absorption bands of hemoglobin and its derivatives

Compound	No of Bands	Situation of Wave Lengths of Absorption Bands in Nanometer (nm)		
Oxyhemoglobin	2	578	542	
Reduced hemoglobin	1	565		
Carboxyhemoglobin	2	569	539	
Sulfhemoglobin	3	618	578	540
Reduced heme	2	607	582	
Cyanmethemoglobin	1	540		

oxygenation of the hemoglobin molecule causes the release of a proton This proton in turn reacts with bicarbonate to liberate CO_2 according to the reaction outlined below

$$HHB + O_2 \rightleftharpoons HbO_2 + H^+ + HCO_3^- \rightleftharpoons H_2CO_3 \rightleftharpoons H_2O + CO_2$$

The reactions take place in the lungs in the direction indicated, in the peripheral tissues the reverse reaction takes place and allows more CO_2 to be carried in venous blood About 60 to 70% of the total CO_2 transport is accounted for by this mechanism Within the red cell a zinc-containing enzyme, carbonic anhydrase, greatly facilitates the transport of CO_2 from the tissue to the lungs by its action in the reversible formation of H_2CO_3

$$H_2O + CO_2 \rightleftharpoons H_2CO_3$$

A second way in which hemoglobin is important in the transport of carbon dioxide is through the formation of carbamino complexes

$$RNH_2 + CO_2 \rightarrow RNHCOOH$$

Deoxyhemoglobin binds more CO_2 as a carbamino compound than does oxyhemoglobin, and thus, the shuttle of the hemoglobin between oxyhemoglobin and reduced hemoglobin provides an additional mechanism for the transport of CO_2

Carbon Monoxide. Carbon monoxide unites with hemoglobin in the same proportions as does oxygen, competing successfully with the latter for hemoglobin and displacing oxygen volume for volume to form carboxyhemoglobin Unlike oxygen, however, the combination of carbon monoxide with hemoglobin results in a stable compound which can be disrupted only with the greatest difficulty The

much greater avidity (between 200 and 250 times) which hemoglobin shows for carbon monoxide renders the gas highly dangerous Nitric oxide gas also has a strong affinity for hemoglobin and forms a stable compound with it As with methemoglobin and carboxyhemoglobin, in the presence of nitric oxide-hemoglobin there is an increased affinity for oxygen by hemoglobin (Drabkin 1950)

Hydrogen Sulfide. Sulfhemoglobin a green compound resulting from the reaction of ferrohemoglobin with hydrogen sulfide, ordinarily occurs only when blood undergoes putrefaction Certain drugs, however, notably acetanilid and phenacetin, sensitize hemoglobin so that it combines more readily with hydrogen sulfide Small quantities of the gas that are absorbed from the alimentary canal may then cause sulfhemoglobin to reach a relatively high concentration in the circulation, resulting in a bluish or mauve tint to the skin So-called *enterogenous cyanosis* occurs when sulfhemoglobin amounts to 3 to 5 g /100 ml of blood

OTHER HEME COMPOUNDS

Cytochrome. Cytochromes are heme compounds that are widely distributed in the tissues of plants and animals, where they play a fundamental role in cellular respiration Oxidation of one of the cytochromes by molecular oxygen ultimately results, after a series of coupled oxidation-reduction reactions, in oxidation of a substrate thereby making energy available to the organism The function of hemoglobin by contrast is to deliver oxygen to the cells so that these reactions will take place

Myoglobin Myoglobin the pigment of muscle resembles hemoglobin in its function It is the simplest example of a respiratory heme protein, since the molecule consists of but one polypeptide chain and one heme group It has a molecular weight of about 17 000, the iron content is 0 32% It acts as an oxygen reservoir within the muscle fiber Since it has a higher capability for oxygen binding than does hemoglobin, it becomes oxygenated more rapidly than hemoglobin at any given Po_2 (Antonini, 1965)

Porphyria Disorders of porphyrin metabolism, in which abnormal porphyrins or large quantities of normal porphyrins are excreted, are collectively known as porphyrias The inherited (primary) abnormalities of porphyrin metabolism are customarily classified as erythropoietic porphyria, acute intermittent (hepatic) porphyria, porphyria cutanea tarda, and mixed (hepatic-cutaneous) porphyria The term porphyrinuria refers simply to the increased excretion of porphyrins in the urine, whether this be on a primary or secondary (acquired) basis Detailed descriptions of these disorders may be found in cited references (Harris, 1970)

5 / TRANSFUSIONS: THE BLOOD GROUPS

Blood transfusion is required to maintain adequate oxygen transport in anemia and to restore the blood volume after hemorrhage from trauma or disease Different fluids have been used for these purposes physiological saline, solutions containing compounds of high molecular weight such as polyvinylpyrrolidone and dextran, plasma fractions, concentrated red cells, whole blood, and washed red cells Of these, only adequately preserved red cells can restore oxygen transport Substances of small molecular weight maintain the blood volume for only a short time, whereas solutes of large molecular weight remain in the circulation for a longer time and may produce toxic reactions

CONSIDERATIONS FOR BLOOD TRANSFUSION

Transfusion techniques are relatively safe providing certain precautions are taken Careful selection of donors is important to avoid transmission of homologous serum jaundice, malaria, and syphilis (Blood is now tested for the presence of the Australian antigen, although its exact nature has not been established, its presence is closely associated with serum hepatitis) The blood must be sterile The blood transfusion is immunologically satisfactory when neither the recipient nor the donor blood cells are destroyed by antibodies in donor or recipient plasma, and when the transfused cells do not stimulate antibody formation in the recipient The major consideration in blood transfusion is the determination of blood groups of the donor and the recipient and selection of compatible donor blood Landsteiner (see 1936) at the turn of the century, described the ABO blood group system along with many of the basic facts regarding the immunological reactions of the erythrocytes. Usually, a blood transfusion will be clinically acceptable (successful) if ABO and Rh compatible blood is used However, there exists a potential for immunization from other blood group antigens (Mollison, 1967, Schmidt, 1969, Strumia et al , 1963)

Blood Groups

Lansteiner (1936) and others classified blood into four groups, A, B, AB, and O The classification is based on the antigenic nature of the red cells Other blood groups were later discovered, including M, N, P, and many more

Most erythrocyte antigens are not clinically significant except in rare cases In some instances, however, they may be responsible for transfusion reactions and hemolytic disease of the newborn (see below) Table 4 7 shows most of the human blood group systems and some of the related antigens within each system Table 4 8 shows the incidence of various antigens in the general American population Blood group antigens are in the cell membrane

Certain antigens are characteristic of ethnic groups—e g , the D (Rh_0) antigen is present in approximately 85% of the Caucasian population, but is present in 95% of the Negro and 100% of the Oriental and the American Indian population A complete discussion of all antigenic systems is beyond the scope of this chapter Emphasis will be placed on systems of greatest clinical importance

THE ABO SYSTEM

Determination of Groups When antigen A is present on cells, they react with serum that contains corresponding antibodies called anti-A, so that large numbers of cells clump together (agglutinate) This reaction is the same as that between any antigen and its antibody Similarly, cells with antigen B will react with anti-B antibodies Cells containing A and B antigens are agglutinated by either antibody, cells without A or B antigens (group O

Table 4 7
Blood group antigens in man

System	Antigens Detected by: Positive reaction with specific antibody	Antigens Detected by: Positive reaction with one antibody negative with another
A_1A_2BO	A_1B, H^b	A_2, A_3, A_x and other A and B variants
MNSs	M, N, S, s, U, M^g, M_1, Tm, M^k, Hu, He, $M1^a$, Vw(Gr), Mur, Hil, Vr, Ri^a, St^a, Mt^a Cl^a, Ny^a, Sul, Sj	M_2, N_2, M^c M^a M_3, S_2
P	P_1, pk, $Luke^b$	P_2
Rh	D, C, c, C^w C^x, E, e, c^e(VS), E^w, G, ce, ce^s (V), Ce, CE, cE, D^w, E^T, LW^b	D^u, C^u, E^u, many other variant forms of D, C c and e
Lutheran	Lu^a, Ku^b	
Kell	K, k, Kp^a, Kp^b, Js^a, Js^b	
Lewis	Le^a, Le^b	
Duffy	Fy^a, Fy^b	
Kidd	Jk^a, Jk^b	
Diego	Di^a, Di^b	
Yt	Yt^a, Yt^b	
I	I, i	
Xg	Xg^a	
Dombrock	Do^a	
Very frequent antigens	Vel, Ge, Lan, Co^a, Gy^a, At^a, and several unpublished examples	
Very infrequent antigens	Levay, Wr^a, Be^a, By, Sw^a, Good, Bi, Tr^a, Wb, Bp^a, Rd, Ls^a, Box, Or, Ht^a, Gf, Wu, and many unpublished examples	
Other antigens	Au^a (Auberger) Sm and Bu^a ("allelic") Bg, Cs^a, Chido, Sd^a, Ul^a, Go^a	

From Race and Sanger, 1968
[a] Recognizable only in favorable genotypes
[b] A genetically independent part of the system

Table 4 8
Occurrence of blood group antigens in the general American population

Blood Group System	Antigen	Positive	Negative
ABO	A	43	57
	B	12	88
MNSs	M	78	22
	N	72	28
	S	55	45
	s	89	11
P	P_1	79	21
Rh	C	68	32
	c	81	19
	D	85	15
	E	29	71
	e	98	2
Lutheran	Lu^a	8	92
	Lu^b	99 8	0 2
Kell	K	9	91
	k	99 8	0 2
Lewis	Le^a	22	78
	Le^b	72	28
Duffy	Fy^a	66	34
	Fy^b	83	17
Kidd	Jk^a	75	25
	Jk^b	75	25

cells) do not agglutinate with either antibody These findings led to the formulation of the so-called "Landsteiner's law," namely that in the plasma of persons with cell group A, anti-B antibodies are present, in the plasma of persons with cell group B, there are anti-A antibodies, group AB blood contains neither anti-A nor anti-B antibodies, and group O blood contains both anti-A and anti-B antibodies (Table 4 9) Obviously, if the serum from a person of group A is mixed with the cells of a person of group B, agglutination will occur If the same serum is added to group O cells, no agglutination will occur The antibodies of the ABO system are naturally present in human plasma (or serum) but the mechanism of their production is uncertain

Blood group antibodies (isoantibodies) are present in the γ globulin fraction of the plasma By using two antisera, one containing anti-A and the other containing anti-B, it is possible to determine to which of the four groups of the ABO system any sample of blood belongs Agglutination of the cells by only the anti-A serum shows that the cells are of group A, while agglutination only with anti-B serum

Table 4 9
Antigens and antibodies of major blood groups

Group	Cells Contain Antigens	Serum Contains Antibodies
O	O	Anti-A and anti-B
A	A	Anti-B
B	B	Anti-A
AB	A and B	Neither

indicates that the cells are of group B If agglutination occurs with both antisera, the cells are of group AB, and if agglutination does not occur with either antiserum, the cells are of group O

Blood Group Antibodies. Isoantibodies fall into two groups based on sedimentation constants The 19S globulins have a molecular weight of about 1,000,000 They usually agglutinate red cells suspended in saline The 7S globulins have a molecular weight of about 160,000 They agglutinate red cells only in a modified medium containing a high protein concentration or after treatment with certain enzymes Presence of the 7S globulin on the red cell surface can also be detected by the antiglobulin test of Coombs, Mourant, and Race (see next paragraph) The 7S globulins pass the placental barrier and may cause hemolytic disease of the newborn Agglutination or, in some cases when complement is present, hemolysis reveals the presence of antibodies on the red cell

Antiglobulin (Coombs) Test. The antiglobulin test has been of great practical importance in the field of blood transfusion and its application has enhanced our knowledge of blood group systems (Coombs et al , 1945, Bove, 1967, Mollison, 1967) Antihuman globulin serum (Coombs serum) contains antibodies against human γ globulin, and therefore, it reacts with all human antibodies Red cells coated with 7S γ globulin are not agglutinated because 7S γ globulin is an "incomplete" antibody However, upon the addition of Coombs serum to the reaction medium, agglutination occurs The action of the antiglobulin antibody may be visualized as a link that binds the blood group antibodies together, these in turn are bound to red cells which are thus agglutinated When the addition of antiglobulin serum to a sample of washed red cells causes agglutination, it indicates the presence of 7S γ globulin on the red cell surface (see Fig 4 22) When the test is properly conducted, the inference that the globulin is a specific antibody is permissible (Nonantibody-caused, Coombs serum-induced agglutination is possible) This test, called the *direct Coombs test*, is used to detect the presence of antibodies already bound to red blood cells, as in hemolytic disease of the newborn or "autoimmune" acquired hemolytic anemia In contrast, in the *indirect Coombs test*, the antibody coating of red cells is induced in vitro, using a known though incomplete antibody, to aid in blood typing or for purposes of antibody identification (using red cells of known antigenic composition) (see Fig 4 22)

Subgroups of A. The existence of subgroups of A, called A_1 and A_2, reflects differences both qualitative and quantitative in the expression of the A antigen on the surface of the red cells This terminology is somewhat confusing in that cells of group A_1 contain both A and A_1 antigens, and cells of group A_2 contain only A antigen The use of an antiserum specific for A_1 permits this distinction Group AB may be subdivided into A_1B and A_2B

Inheritance of ABO Blood Groups. The gene O is an amorph It has no recognizable specific gene product It is now believed that there are 4 alleles, A_1, A_2, B, and O, and that

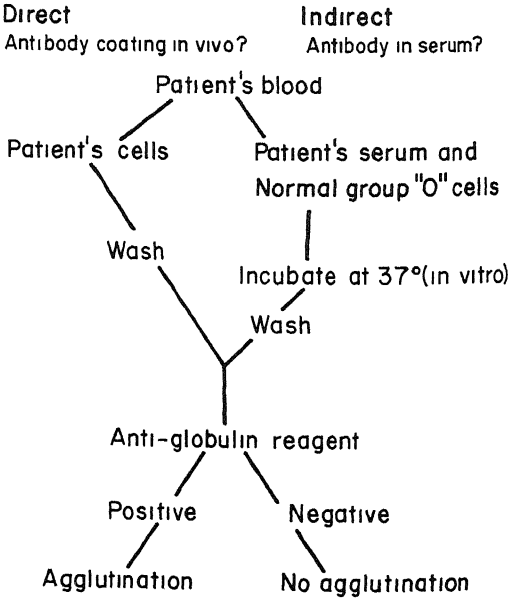


Fig 4 22 Direct and indirect Coombs test (From Bove 1967)

these yield 10 genotypes with 5 distinguishable phenotypes (Table 4 10)

Other Sources of Blood Group Antigens. The antigens are typically a part of the red cell membrane, but many other body tissues and fluids can usually but not always produce serological reactions with the A and B antigens The antigens on red cells have glycolipids, those in body fluids are polysaccharides The antigenic sites are probably chemically identical Persons whose saliva, gastric juices, etc contain appreciable amounts of ABO antigens are classified as *secretors* as opposed to *nonsecretors* *Secretors* of ABO do not have the Lewis antigen (Le^a), whereas most *nonsecretors* have this antigen on their erythrocytes and in their body fluids

The A and B substances present in body fluids have a low nitrogen content (compared with protein) and high concentrations of reducing sugar and hexosamines Antigenic specificity lies in the sugar component, e g group A specificity is associated with α-D-N-acetylgalactosamine, whereas B specificity is associated with α-D-galactose

M, N AND P ANTIGENS

Landsteiner and Levine (1928) demonstrated the existence of additional antigens, M, N, and P By using anti-M and anti-N sera, cells may be classified as either M, N, or MN M and N are poor antigens in man, and anti-M and anti-N are rarely found in human sera Antigen P is of little clinical importance

Table 4 10
Genes of major blood groups

Genotypes	Phenotypes
A_1A_2, A_1A_2, A_1O	A_1
A_2A_2, A_2O	A_2
BB, BO	B
A_1B, A_2B	AB
00	0

THE RH BLOOD GROUP SYSTEM

Landsteiner and Wiener (1940, 1941) immunized rabbits and guinea pigs with the red cells of a Rhesus monkey The resulting antiserum produced agglutination in a large portion of the human red cells studied, the responsible human blood antigen was called Rh Individuals with the antigen are Rh-positive About 85% of Caucasians are positive and 15% are Rh-negative There are several Rh antigens "Rh-positive" and "Rh-negative" refer to the presence or absence of the most important, the D antigen

There is no known antigenic stimulus that accounts for the production of antibodies of the ABO system In contrast, antibodies to the Rh antigen are almost always produced by exposure to the "foreign" Rh antigen, but only by those individuals who lack the corresponding antigen Isoimmunization occurs in two ways *1)* a transfusion of Rh-positive red cells to an Rh-negative recipient, or *2)* the transplacental displacement of Rh-positive cells of the fetus into an Rh-negative mother

Hemolytic Disease of the Newborn (HDN). Hemolytic disease of the newborn or *erythroblastosis fetalis* is a serious complication of pregnancy affecting the fetus in utero The child may be born dead or with hemolytic anemia High concentrations of bilirubin, a degradation product of hemoglobin, may damage brain cells and produce a condition called kernicterus

The sequence of events causing hemolytic disease of the newborn is as follows an RH-negative mother conceives from an Rh-positive father, and an Rh-positive fetus results In some manner, Rh-positive fetal cells enter the maternal circulation and stimulate production of anti-Rh antibodies The antibodies return to the fetus attacking fetal red cells in proportion to their avidity and concentration HDN usually is not produced in first pregnancy, several Rh-positive pregnancies may take place before HDN develops During subsequent pregnancies the antibodies increase in potency and cross the placental barrier into the fetal circulation They react with the fetal Rh-positive cells and destroy them A transfusion of Rh-positive cells to an Rh-negative woman years prior to a pregnancy increases the probability of HDN with later pregnancies

Hemolytic disease of the newborn is treated by "exchange transfusion," i e , by slowly re-

moving blood from the affected infant and simultaneously replacing it with Rh-negative blood By this procedure, the anti-Rh antibodies are removed along with bilirubin Within a few weeks the transfused cells are removed from the circulation and are replaced by the infant's own Rh-positive cells Recently, in an attempt to reduce the number of stillbirths due to HDN, Rh-negative blood has been infused into the abdominal cavity of the fetus in utero The transfused cells are absorbed into the fetal circulation, where they circulate without being attacked by the anti-Rh antibodies

In addition to the various forms of therapy for the established disease, an approach to prophylaxis is now possible through the application of the principle of "antibody-mediated immunosuppression" Pollack et al (1969) have administered Rh-positive antibodies to previously nonimmunized Rh-negative women following delivery The passively administered Rh antibodies prevent Rh antibody formation (Pollack et al , 1969) This principle of prevention may be applicable in other conditions

Hemolytic disease may also result from ABO incompatibility between mother and fetus, but this is usually milder than that caused by Rh incompatibility In rare cases other antibodies are responsible

Nomenclature. Shortly after the discovery of the Rh antigen a number of related antigens were found Two systems of nomenclature, based on slightly different concepts of the genetics of the Rh antigen, have been widely used to describe them (Race and Sanger, 1968) In recent years as many variants of the Rh antigen have been found, the terminology to describe these variants has become increasingly complex (Wiener and Wexler, 1958) A third nomenclature, using numbers, has been proposed (Rosenfield et al , 1962) The reader is referred to the paper of Wiener and Wexler (1958) for a discussion of the Rh nomenclature, to the paper of Fisher (see Race and Sanger, 1968) for a discussion of the CDE nomenclature, and to publications by Rosenfield and collaborators (1962) for a discussion of numerical terminology

A complete discussion of all the Rh variants is not possible in this text, but a brief discussion of some of the more common ones follows Table 4 11 shows the reactions of the Rh antigens with the more commonly available antisera

According to Race and Sanger (1968) there are at least three sets of antigens C and c, D and d, E and e The corresponding antibodies are anti-C, anti-c, anti-D, anti-E, and anti-e Anti-d has not yet been found C antigen corresponds to Wiener's rh′, D antigen to Rh_0, and E antigen to rh″ The antigen combination CDe corresponds to Wiener's Rh_1 An Rh-negative type is shown as cde It should be noted that the use of lower case letters implies not only the absence of one of the Rh antigens, but also the presence of its reciprocal antigen

D, or Rh_0, is by far the strongest immunogenic antigen In European populations most Rh-positive cells are CDe, or Rh_1

Reciprocality and Absence of Rh Antigens. There seems to be a reciprocal relation between blood groups in the Rh system, i e , the absence of a certain antigen implies the presence of another (reciprocal) For example, if there is no C antigen, then the c antigen

Table 4 11
Reaction of Rh antigens with antisera

Cells		Antiserum				
Antigen combination (Wiener agglutinogen symbol)	Gene symbol	Anti-C (anti-rh′)	Anti-D (anti-Rh_0)	Anti-E (anti-rh″)	Anti-c (anti-hr′)	Anti-e (anti-hr″)
cDe (Rh_0)	R^o	−	+	−	+	+
Cde (Rh′)	r′	+	−	−	−	+
cdE (Rh″)	r″	−	−	+	+	−
CDe (Rh_1)	R^1	+	+	−	−	+
cDE (Rh_2)	R^2	−	+	+	+	−
CDE (Rh_z)	R^z	+	+	+	−	−
CdE (Rh_y)	r^y	+	−	+	−	−
cde (rh)	r	−	−	−	+	+

must be present, or the absence of rh' implies the presence of hr' The absence of the E (rh') antigen implies the presence of e (hr') Anti-c and anti-e sera may be used to demonstrate the existence of the alternative antigens One might expect that a d (Hr_0) antigen exists as an alternative to the D (Hr) antigen, however its existence has never been demonstrated

Very rarely are persons encountered in whom the Rh antigens are absent but several people who have no detectable representatives of any of the Rh antigens i e (- - -) have been reported This is called an *Rh null*

OTHER BLOOD GROUPS

Other blood groups include Kell (K) and Cellano (k) (which bear a reciprocal relationship) and Duffy (Fy^a, Fy^b), Diego (Di^a), Lewis (Le^a, Le^b), Lutheran (Lu^a, Lu^b), and Kidd (Jk^a, Jk^b) (Table 4 7) Some 'private" blood group antigens are infrequently seen in the general population, and positive reactions occur only in members of a particular family

Transfusion of Blood

EFFECTS OF INCOMPATIBLE TRANSFUSIONS

When blood from a person of group A is transfused into a recipient of group B, there are two possible ways by which agglutination of red cells may occur the anti-B antibodies of the donor may agglutinate the B cells of the recipient, or the anti-A antibodies of the recipient may agglutinate the A cells of the donor Generally the donor antibodies are so well diluted in the recipient's plasma that this type of agglutination is rare The antibodies in the recipient's plasma, on the other hand, are only slightly diluted by the small amount of blood usually given, and readily react with the cells of the donor (Mollison, 1967, Schmidt, 1969)

Reactions can ordinarily be avoided by transfusing group O blood, since these cells have neither A nor B antigens to be agglutinated by the recipient's anti-A or anti-B antibodies The anti-A and anti-B antibodies present in O blood usually are diluted enough so there is no effect on the recipient's cells, and the term "universal donor" has been applied therefore to persons with group O blood Persons with group AB blood are sometimes called "universal recipients " These terms are misleading because some people with group O blood have high concentrations of anti-A or anti-B antibodies of strong hemolytic nature which despite dilution may destroy recipient cells Such reactions can be avoided by using washed red cells

If blood of the "wrong (incompatible) ABO blood group is transfused a hemolytic transfusion reaction is probable Destruction of red cells produces hemoglobinemia and hemoglobinuria Chills fever, shock, renal insufficiency and death may ensue

Routine crossmatching is performed only for the ABO and Rh systems therefore, isosensitization may develop in recipients who receive donor red cells that contain antigens which the recipient lacks (Table 4 11)

STORED BLOOD AND BLOOD COMPONENTS

Fresh blood is rarely transfused today Blood is usually collected and stored in the liquid or frozen state Liquid whole blood contains many substances that may cause adverse reactions, therefore, its routine clinical use has been repeatedly criticized Often it is recommended that only the required blood component be used Today whole blood after collection can be fractionated into its various components concentrated red cells, platelet-rich plasma or platelet concentrates, fresh frozen plasma, antihemophilic globulin or cryoprecipitate, albumin, and various γ globulin fractions Liquid-preservation, the principal method of storing red cells, extends the shelf-life of whole blood only to 3 to 4 weeks Freeze-preservation extends the shelf-life of red cells many months and perhaps indefinitely (Bishop and Surgenor, 1964, Valeri, 1970)

Preservation Injury. Red cell preservation is satisfactory when the transfused red cells circulate and function immediately after transfusion Preservation procedures involve use of a liquid medium with storage at 4° C, a freeze-preservation technique using various cryoprotective agents and storage at either −80° C or −150° C, or a combination of the liquid- and freeze-preservation approaches Blood for freezing is collected in one of several liquid preservatives, acid-citrate-dextrose (ACD), citrate-phosphate-dextrose (CPD), ACD plus adenine, or CPD plus adenine with or without inosine, it is then frozen by one of a number of

procedures, stored at either −80° C or −150° C, thawed, and washed prior to transfusion Either previously frozen glycerolized red cells or nonfrozen red cells stored as concentrated red cells or as whole blood can be washed and resuspended in a nonplasma-crystalloid medium and stored at 4° C up to 24 hr prior to transfusion These manipulations, however, produce "preservation injury" which manifests itself in the form of severely and irreversibly damaged red cells that are removed from the recipient's circulation at an accelerated rate during the transfusion and the 24-hr post-transfusion period Figure 4 23 shows that "preservation injury" affects only the 24-hr post-transfusion survival of preserved red cells Preservation injury is a composite effect that reflects a combination of lesions produced by the following conditions collection, storage in the liquid state prior to freezing, storage in the frozen state, freeze-thaw per se, red cell washing, and storage after thawing in the liquid state with or without post-thaw washing or any other manipulation "Preservation injury" is to be distinguished from immunological incompatibility between the donor red cells and the recipient

With the in vitro compatibility tests currently in use (saline, albumin, and antiglobulin crossmatching), red cell incompatibility can, with rare exception, be ruled out as the cause of the rapid removal of nonviable red cells during the 24-hr post-transfusion period There is no in vitro measurement, however, that can predict the 24-hr post-transfusion survival value of preserved red cells The rapid removal (phagocytosis) of nonviable but compatible red cells suggests a defect in the red cell membrane recognized by the recipient's reticuloendothelial system The rate of removal of damaged red cells during transfusion is affected by the recipient's health seriously ill patients have an ineffective removal mechanism, whereas otherwise healthy recipients with accidental injuries promptly remove irreversibly damaged red cells Furthermore, the method of preservation is a determining factor in the mechanism of removal of the nonviable red cells Liquid-preservation procedures (i e , collection and storage in either acid-citrate-dextrose (ACD), National Institutes of Health, Formula A or citrate-phosphate-dextrose (CPD) produce nonviable compatible red cells that are removed by extravascular sequestration without intravascular liberation of extracorpuscular hemoglobin Some freeze-preservation procedures may produce compatible but nonviable red cells that are removed with the intravascular liberation of hemoglobin from the cells (Valeri, 1970)

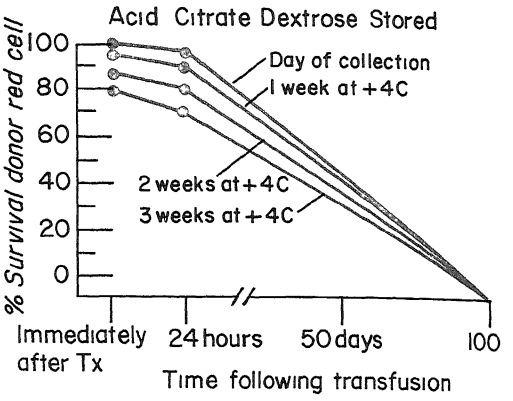


Fig 4 23 The relationship between the fraction of red cells surviving after transfusion to storage time in ACD

LIFESPAN OF PRESERVED RED CELLS

Nonviable red cells are removed during the 24 hr following transfusion Figure 4 23 shows that red cells that remain in the circulation have a *potential* for normal long-term survival In vivo restoration of reversibly damaged liquid-stored or previously frozen red cells has also been demonstrated A rapid decrease in donor red cell sodium level and slow increase in donor red cell potassium concentration, together with a change in red cell osmotic fragility towards normal indicate that preserved red cells are restored during circulation in the recipient A satisfactory preservation procedure maintains normal intracellular sodium and potassium levels and normal osmotic fragility until the time of transfusion Viable transfused red cells have a *potential* for normal lifespan, however, the lifespan may be adversely affected by the recipient's intravascular environment, e g , certain immunological, chemical, and mechanical factors Post-transfusion mean lifespan of those red cells that are viable is not related to the preservation technique nor to the length of storage in the liquid or frozen state

FUNCTION OF PRESERVED RED CELLS

"Preservation injury" may markedly affect oxygen transport by an increased affinity of

the hemoglobin for oxygen In blood collected in ACD the increased affinity for oxygen is correlated with a decrease in the 2 3-diphospho-D-glyceric acid (DPG) concentration of the red cells (see chap 4) A slight decrease in the red cell adenosine triphosphate (ATP) level also occurs After transfusion there is in vivo restoration of the oxyhemoglobin dissociation curve usually within 24 hr and an increased 2,3 DPG concentration The quantity and quality of the preserved red cells and the physical condition of the recipient determine the rate of in vivo restoration of the red cell oxygen dissociation curve towards normal

A relation has been observed between the oxyhemoglobin dissociation curve and the levels of red cell organic phosphate compounds (ATP and 2,3 DPG) in blood preserved in acid-citrate-dextrose (ACD), ACD plus adenine, citrate-phosphate-dextrose (CPD), or CPD plus adenine The addition of pyruvate, inosine, and phosphate to these media replenishes the levels of red cell 2,3 DPG and ATP and restores toward normal the oxyhemoglobin dissociation curve

After blood preserved for 15 days is transfused into anemic recipients there is progressive in vivo restoration towards normal of 2,3 DPG, ATP, potassium and sodium ion concentrations Within 8 hr of the transfusion a restoration of at least 25% of the final level of red cell 2,3 DPG occurs, and within 24 hr at least 50% of the final level is achieved Thereafter, the increase in donor red cell 2,3 DPG is curvilinear and reaches a plateau usually within 9 to 11 days after transfusion The ATP level of the transfused red cells rises rapidly during the first 24 hr , this rise is associated with a rapid fall in donor red cell sodium concentration Donor red cell potassium levels rise very slowly, and the increase in cellular potassium concentration appears to be related to the absolute level of red cell 2,3 DPG (Valeri, 1970)

Current methods of freeze-preservation of human red cells maintain normal levels of the organic phosphates It is now possible to combine liquid- and freeze-preservation methods to provide functioning red cells and help to resolve problems of imbalance of supply and demand for human red cells One can collect blood in CPD-adenine and store it for at least 2 weeks at 4° C If the blood is not used during this initial period, pyruvate, inosine, and phosphate can then be added to restore the 2,3 DPG and ATP pools, after which the red cells can be frozen using one of the glycerol procedures Current methods of freeze-preservation of human red cells use glycerol an intracellular cryoprotective substance therefore post-thaw washing is essential This reduces the amount of adenine and inosine that might otherwise produce untoward effects after multiple-unit transfusions However, freezing human red cells immediately upon collection obviates the need for purine nucleoside supplementation and permits separation of whole blood into its components Each blood component can then be stored separately using the appropriate technique for optimal preservation

Preservation and reconstitution techniques must also minimize the hemolysis that may occur Haptoglobin of plasma binds hemoglobin If, in the course of destruction of nonviable red cells in vivo, the haptoglobin-binding capacity of plasma is exceeded there will be free extracorpuscular hemoglobin Bound hemoglobin is cleared by the reticuloendothelial system Free extracorpuscular hemoglobin, either infused with the supernatant of the blood unit or liberated in vivo from the removed nonviable red cells, may saturate the hemoglobin-binding capacity of the recipient and then if the tubular reabsorptive capacity of the kidney is exceeded, the excess free plasma hemoglobin will be excreted in the urine Hemoglobinuria can be tolerated by normotensive, normovolemic subjects with normal renal function It may precipitate acute renal insufficiency, however if associated with hypotension, an increased secretion of the antidiuretic and renin-angiotensin hormones, dehydration, acidosis, or hemorrhage

If no intravascular hemolysis occurs, a limiting factor in the clinical acceptability of preserved red cells is the amount of infused free supernatant hemoglobin This does not mean necessarily that hemoglobin is toxic, the damage may be produced by red cell stroma liberated simultaneously with the hemoglobin during hemolysis

Some degree of in vitro hemolysis occurs with all types of preservation, resulting in the liberation of hemoglobin, potassium, enzymes, purine nucleotides, and the red cell membrane The presence of a significant amount of red cell stroma (ghosts and fragments) may precipitate renal insufficiency by sludging in the microcirculation or by activating dissemi-

nated intravascular clotting. Intravenous potassium is potentially myocardiotoxic; thus, the amounts of these substances present in the preserved blood must be kept within safe limits.

Washing frozen or nonfrozen red cells effectively removes in part *1*) white cells and platelets; *2*) isoagglutinins; *3*) the anticoagulant preservative and red cell metabolites; and *4*) plasma factors, protein and nonprotein (allergens, etc.). The clinical indications for the transfusion of washed red cells (liquid-preserved or previously frozen) are:

1) A history of febrile transfusion reaction, or urticarial and anaphylactoid transfusion reaction—attributable to recipient antibodies.

2) The need to avoid overloading the circulation—caused by vasoconstrictor plasma substances in preserved blood and excess sodium.

3) Preparation for allografting procedures—because removal of platelets and leukocytes will minimize probability of presensitization to donor organ.

4) Desire to reduce the probability of a graft vs. host reaction in recipients with immunological deficiency diseases.

5) Intention to decrease the probability of post-transfusion hepatitis by washing out the virus.

HEMOLYSIS AND SUSPENSION STABILITY OF THE BLOOD

Hemolysis in vitro results in escape of hemoglobin from the red cell into the surrounding medium whereas hemolysis in vivo results in a reduction of the lifespan of red cells There are two mechanisms by which red cells may be destroyed in vivo *1)* disruption or dissolution within the general circulation by the action of immune bacterial, or chemical lysis resulting in *intravascular hemolysis* and liberation of hemoglobin into plasma, and *2)* the more common sequestration of red cells in cells of the reticuloendothelial system particularly in the spleen and liver, where the cellular hemoglobin is degraded producing the clinical picture of *extravascular hemolysis* Two primary mechanisms for the red cell sequestration have been proposed *a)* by a surface adherence of the cells to the reticuloendothelial cells that line the blood vessels, followed by phagocytosis (essentially a chemical or electrostatic process), or *b)* by simple filtration (essentially a physical process) (Jandl et al , 1961, Rifkind, 1966) (See chap 3, spleen)

Hemolysis in Vitro

Hemolysis in vitro liberates red cell constituents (e g , hemoglobin potassium purine nucleotides, and the red cell membrane-stroma) into the supernatant fluid The erythrocyte membrane consisting of lipids, ions, proteins, and water is altered during hemolysis The membrane lipids include phospholipids, neutral lipids (free and esterified cholesterol), and glycolipids The erythrocyte membrane composition of cholesterol and the lysolecithin-lecithin ratio are influenced by the plasma enzyme levels of lecithin-cholesterol acyltransferase and phosphatidases (Cooper and Jandl, 1968, Ferber et al , 1970, Jeannet and Hassig, 1964) The protein composition of the red cell membrane, especially the membrane sulfhydryl activity, is related to hemolysis both in vitro and in vivo (Jacob and Jandl, 1962) In the autohemolysis test sterile defibrinated blood is incubated at 37° C for 24 and 48 hr and the amount of spontaneous hemolysis is determined The autohemolysis test is a nonspecific test and reflects the red cell metabolic and physical-structural states (Selwyn and Dacie 1954)

The permeability of the red cell membrane is quite different from that of the capillary membrane Under physiological conditions the membrane of human erythrocytes is impermeable to hemoglobin to the plasma proteins and to the organic phosphate compounds but it does permit the passage of water H^+ NH_4^+, Cl^- HCO_3^- and $PO_3^{\equiv}$ It is permeable also to the Na^+ and K^+ ions, these ions exchange across the membrane and for each there is a specific energy-dependent gradient between the red cells and the plasma Under conditions which injure the red cell membrane, potassium escapes from the cells and the potassium loss is closely related to the hemoglobin loss

OSMOTIC FRAGILITY

The osmotic fragility test, an in vitro test to determine the susceptibility of a red cell population to a hypotonic gradient, reflects the surface-to-volume ratio, intracellular osmotic pressure, and membrane permeability of the red cell population Osmotic fragility in vitro is influenced by *1)* the suspending medium whose pH and tonicity are controlled, *2)* the total number of intracellular osmotically active constituents which determine cell volume in any given external environment, and *3)* the critical hemolytic volume (the mean red cell volume for the population tolerated before total hemolysis occurs), which varies with quantitative and qualitative factors associated with the composition of membrane lipid and protein (Weed and Bowdler, 1966) Normal red cells swell when placed in 0 6% saline solution (osmolarity is 200 mosmole) with a controlled pH and temperature, although the mean red

cell volume increases, the intracellular hemoglobin does not escape (Normal plasma has an osmolarity of 300 mosmoles/liter, equivalent to a 0 9% NaCl solution) Hemolysis begins at a salt concentration of about 0 42% when the red cell volume increases to about 145% of normal At 0 35% saline solution, the red cells are fully hemolyzed, the red cell volume is about 165% of normal (the critical hemolytic volume) The osmotic fragility test detects red cells with increased fragility in patients with intrinsic red cell abnormality (i e , hereditary spherocytosis) and with acquired spherocytosis (antibody-coated red cells) Increased osmotic fragility may indicate a decreased surface-to-volume ratio of red cells (resulting from loss of red cell membrane cholesterol alone, or loss of red cell membrane fragments containing cholesterol and phospholipid) and an increase in membrane permeability to the sodium ion (hereditary spherocytosis) (Cooper and Jandl, 1969, Jacob and Jandl, 1964, Jacob and Karnovsky, 1967, Weed and Reed, 1966)

Decreased osmotic fragility of red cells is usually associated with iron deficiency anemia, hereditary hemolytic disorders of thalassemia and sickle cell anemia, and liver disease with obstructive jaundice There is an increased surface-to-volume ratio, with an increased membrane cholesterol content secondary to decreased plasma lecithin-cholesterol acyltransferase activity, in patients with elevated levels of bile salts, and there is a decreased intracellular osmotic pressure from a decrease in the amount of hemoglobin in the red cells of patients with iron deficiency anemia (Cooper and Jandl, 1968, Wintrobe, 1967) Osmotic hemolysis does not occur usually in vivo Nevertheless, the in vitro test of osmotic fragility has been helpful in the evaluation of red cell abnormalities

ACID HEMOLYSIS

Acid hemolysis is used in vitro to determine the susceptibility of red cells to a hydrogen ion gradient When the temperature and the osmotic pressure of the environment are maintained constant, this test can be used to diagnose an acquired defect of red cells called paroxysmal noctural hemoglobinuria The sensitivity of red cells to acid hemolysis can be detected by exposing them to an acid environment, to thrombin, or to a low ionic medium ("sugar-water test") (Hartmann and Jenkins, 1966)

IN VITRO CHEMICAL HEMOLYSIS

Ether, chloroform, benzene, and alcohol hemolyze by dissolving the lipid constituents of the red cell membrane Surfactant substances, such as bile salts, alkalis, and saponin, also cause hemolysis in vitro These phenomena, however, are of no clinical significance

MECHANICAL FRAGILITY FILTERABILITY REVERSIBLE DEFORMABILITY, AND AGGLUTINABILITY OF RED CELLS

Because red cells are subjected to mechanical trauma during normal circulation, in vitro tests capable of simulating the in vivo state are important A simple test of limited value has been developed to quantitate the mechanical fragility of red cells by standardized rotation of blood in a flask with glass beads Teitel (1965) has studied the disc-sphere transformation and plasticity alteration of red blood cells by observing the filterability of washed red cells through filter paper under controlled conditions, in a variety of hematological conditions the results with this in vitro test correlate with the in vivo survival data In addition, reversible deformability of red cells has been determined from the mean pressure required to deform the cells sufficiently to cause them to enter and traverse a micropipette analogue of the microcirculation (LaCelle, 1969, Weed, 1970) The normal biconcave disc-shape of red cells with relatively higher surface area to volume ratio is more deformable than the spherical form of red cells The mean deformability of preserved red cells at specific time intervals correlates with their post-transfusion survival values (Haradin et al , 1969) Thus, shape and deformability of red cells largely determine their post-transfusion survival (Haradin et al , 1969, Teitel, 1965)

Positively charged polyelectrolytes (polybases) induce agglutination of erythrocytes at very low concentrations (Markikovsky et al , 1966) For example, positively charged polylysine attaches to N-acetylneuraminic acid, the major negatively charged component of the red cell membrane, thus, linking cells and producing agglutination The observation that "old red cells" have a reduced negative charge and are agglutinated more easily with polylysine than are "young red cells," indicates a correlation between agglutination and the surface charge of the red cells

Hemolysis in Vivo

Hemolysis in vivo is best defined as shortening of the lifespan of erythrocytes With the exception of incompatible transfusions and some rare disorders, intravascular hemolysis with hemoglobinuria is uncommon Clearance of senescent or defective red cells from the blood is usually performed by the macrophages in the reticuloendothelial system (RES) (See chap 3) It is now apparent that RES-mediated erythrocyte destruction involves the sequential operation of at least three component processes *1) Injury*—the development of a critical, prehemolytic red cell lesion mediated by normal red cell aging, by an intracorpuscular biochemical defect, or by an extracorpuscular agent *2) Sequestration*—the selective filtration of injured red cells from the general circulation and their concentration within the vascular spaces of the RES *3)* Red cell *degradation* by enzymatic digestion within the lysosomal vasuoles of the reticuloendothelial macrophages (Rifkind, 1966; Weed and Reed, 1966)

INTRAVASCULAR HEMOLYSIS

When the rate of hemolysis is excessive, as in blood transfusion reactions, in drug-induced hemolysis, and in malaria (blackwater fever), the plasma extracorpuscular hemoglobin cannot be converted into bilirubin as rapidly as it is liberated, and hemoglobinuria may occur Extracorpuscular hemoglobin in plasma combines stoichiometrically with plasma haptoglobin, a specific α_2 globulin to form a hemoglobin-haptoglobin complex which is cleared from the circulation by the reticuloendothelial system When the concentration of plasma extracorpuscular hemoglobin exceeds the hemoglobin-binding capacity and the tubular reabsorption capacity of the kidney, the excess "free" plasma hemoglobin will be filtered and excreted in the urine A port wine, dark brown, or black colored hemoglobin pigment may be produced by acid urine where hemoglobin may be converted into acid hematin and methemoglobin

Paroxysmal (Cold) Hemoglobinuria. The Donath-Landsteiner hemolysis test is pathognomonic for this disease The unique hemolysin responsible for this reaction is effective against all normal erythrocytes (a panhemolysin) and requires the following conditions in the reaction mixture for its full activity *1)* complement, *2)* chilling to fix the antibody to the red cell and *3)* subsequent warming to induce hemolysis (Harris, 1970)

The Donath-Landsteiner hemolysin is usually but not always associated with a positive reaction to the Wassermann test resulting from congenital or acquired syphilis It may be present only transiently but it usually persists for long periods of time It may or may not disappear following antiluetic therapy, and it has been shown to be separate from the Wassermann factors responsible for the various biological tests for syphilis In a few instances, it has been demonstrated for short periods following infectious diseases (measles) The reason for the development of the peculiar antibody in any of the situations in which it has appeared is not known In many instances hemolytic episodes have been produced by the controlled chilling and warming of an extremity in a susceptible patient with the Donath-Landsteiner hemolysin

Paroxysmal Cold Hemoglobinuria Associated with Cold Agglutinin. A second type of paroxysmal cold hemoglobinuria is distinguishable by the presence of an abnormal amount of cold agglutinin Curiously enough, except under special conditions the agglutinin exhibits little or no lytic activity

The clinical picture and the hematological changes noted may follow very closely those described for the Donath-Landsteiner factor The patients are more apt to develop dermal manifestations, Raynaud's phenomenon, frostbite, and even gangrene of the extremities They are less likely to experience hemoglobinuria with each paroxysm and are more likely to develop a chronically depressed hematocrit and hemoglobin concentration associated with reticulocytosis and hemosiderinuria The chronicity is obviously related to exposure and to climate In the Northern Hemisphere the changes may be noted only during the cold winter months Extraordinarily high titers of the agglutinin are associated with the syndrome With the exception of type I-negative erythrocytes all red cells are agglutinated by the cold antibody Since this blood type occurs in less than 1 in 10,000 individuals, this antibody can be considered a panagglutinin (Harris, 1970)

March Hemoglobinuria. Hemolysis in vivo and hemoglobinuria may occur even in healthy persons after strenuous walking or running In certain persons hemoglobinuria may follow relatively mild exercise The pigment in the

blood and urine is hemoglobin and not myoglobin No red cell abnormalities, hemolysins, or autoagglutinins have been demonstrated Hemoglobinuria occurs after hand trauma, as in the karate chop, suggesting that the disorder is related to mechanical trauma On the other hand, hemoglobinuria decreases during training in marathon runners and is aggravated by sway-back posture This hemolytic mechanism is not simple (Gilligan and Blumgart, 1941)

Nocturnal Hemoglobinuria of Marchiafava In this third type of paroxysmal hemoglobinuria, hemoglobin or hemosiderin is passed almost continuously in the urine, but in greatest amounts at night It is accompanied by a severe hemolytic anemia The hemolysis is caused by the susceptibility of red cells to acid, the greater destruction at night is attributed to the accumulation of carbon dioxide during sleep The abnormal cells are hemolyzed when incubated in slightly acidified normal serum or patient's serum (Ham test), but the normal red cells are not destroyed by the patient's serum The defect in the patient's red cells is unknown This syndrome may produce severe renal damage

INFECTION AND HEMOLYSIS

Three distinct uncommon hemolytic syndromes in patients with infection have been described *1*) an acute, often lethal, intravascular hemolysis during blood-stream invasion by hemolysin-producing organisms such as *Clostridium welchii*, *2*) direct infection of red cells by the red-cell-adherent organism *Bartonella bacilliformis*, causing red cell trapping in the spleen and acute hemolytic anemia; and *3*) an acute or chronic hemolytic anemia developing during the convalescent or postconvalescent phase of certain viral or bacterial infections that stimulate autoantibodies against red cells Such hemolysis is usually attributable to cold agglutinins or to cold hemolysins

The venom of certain poisonous snakes, such as the cobra, and the poisons of various stinging insects and spiders cause destruction of red cells Cobra snake venom liberates an enzyme phosphatidase A, which converts lecithin to lysolecithin, a powerful hemolytic and cytolytic substance Since lecithin is present in erythrocytes, plasma, and all cells, the entrance of snake venom into the body stimulates production of this hemolytic substance (Ferber et al., 1970).

Suspension Stability of Blood

IN VITRO

Sedimentation of freshly collected anticoagulated whole blood is governed by the hematocrit, the viscosity, and the composition of the plasma protein (primarily the fibrinogen and the α and β globulin levels) Whereas, albumin inhibits red cell sedimentation, large asymmetric molecules (fibrinogen and high molecular weight dextran) increase the sedimentation rate

There are various methods of measuring the erythrocyte sedimentation rate (ESR) (Dawson, 1960, Hardwicke and Squire, 1952, Wintrobe, 1967) Because the relative quantities of plasma and red cells influence the sedimentation rate, various methods to correct for the anemic state have been devised These methods may give misleading results since aggregation of the red corpuscles and size of the aggregated particles are far more important factors in rapid and slow sedimentation than the degree of anemia Furthermore, differences in the size and hemoglobin content of the red corpuscles also effect the sedimentation rate, and these have been found to be as important as the plasma factors Variations in red cell size (anisocytosis) may interfere with rouleaux formation (pseudoagglutination, normal red cell tendency to cling together in chains), and variations in red cell shape (poikilocytosis) such as are encountered in sickle cell anemia actually prevent the formation of rouleaux Therefore, in sickle cell anemia sedimentation is very slow even when there is marked anemia The ESR is faster in normal females than in normal males It is a nonspecific reaction which may be compared with the body temperature, pulse rate, and leukocyte count for diagnostic purposes (Wintrobe, 1967)

The conventional capillary tube method of measuring blood viscosity yields information applicable only to the circulatory dynamics in arteries and large arterioles The method detects changes in the concentration of red blood cells, and changes in the concentration and nature of the plasma proteins It does not detect aggregation of the blood elements, unless these aggregates approach in size the diameter of the capillary tube (usually 300 μm) used for testing flow characteristics of the blood It seems likely that aggregates of blood elements, if sufficiently large and held together by sufficiently strong force, would occlude the smaller

arterioles, venules, and capillaries These same aggregates would not interfere with passage of blood through larger blood vessels Aggregates which appear to meet these criteria have been observed in vitro in blood that has been stored for transfusions, during hypotensive endotoxin, and anaphylactic shock, after the ingestion of large amounts of fatty foods and after the addition to blood of large molecular weight substances (dextrans gelatins and the biogenic serotonin and adenosine diphosphate A screen filtration pressure method has been designed to measure the adhesiveness and aggregation of blood cells in an attempt to simulate the flow of blood through the microcirculation (Swank et al., 1964)

The stability of preserved red cells (nonfrozen or previously frozen) after washing and storage at 4° C, for 24 hr is influenced by the composition of the resuspension medium Plasma is not a satisfactory medium because of the possible transmission of homologous serum jaundice Although an artificial medium having a pH of about 7 2 and containing albumin, glucose, phosphate, and sodium chloride adequately maintains the stability of the red cells, it is far from ideal as evidenced by the above described in vitro measurements and their post-transfusion survival

IN VIVO

Normal erythrocytes possess remarkable deformability This is a subject of great interest to those interested in the microcirculation, because deformability is important in determining lifespan in vivo Knowledge of cellular deformability also provides insight into the pathophysiology of shortened red cell lifespan in many hemolytic anemias Alterations in the intrinsic physical properties of erythrocytes may also interfere with blood flow through the microcirculation (Merrill, 1969, Wells, 1964)

Within the microcirculation, where the vital exchanges of oxygen and carbon dioxide take place, there are several factors in addition to the intrinsic properties of the red cells themselves which may influence these physiological functions Circulatory dynamics vessel wall integrity and plasma proteins which may alter the flow properties of whole blood must be normal. Whole blood is non-Newtonian in its fluid and flow characteristics When hematocrit is normal, the high viscosity of blood at low shear rates can be attributed primarily to fibrinogen dependent rouleaux formation which is a common occurrence under conditions of stagnation and can be disrupted by increased shearing forces Rouleaux formation is enhanced by fibrinogen and γ globulin and is inhibited by albumin Certain pathological plasma proteins, such as those found in macroglobulinemia or cryoglobulinemia, may increase the blood viscosity and may help occlude the microcirculation

Prosthetic heart valves may produce a hemolytic anemia

In addition, gram-negative sepsis and posttraumatic states may initiate disseminated intravascular coagulation (DIC) with fibrin deposition in the microcirculation particularly of the lung and the kidney This syndrome of microangiopathic hemolytic anemia often appears very suddenly and is recognized by a distinct appearance of erythrocytes which seem to be contracted, distorted, and sometimes triangular ("helmet cells" or "burr cells") (Brain et al., 1962) It is encountered in a number of conditions characterized by necrotic lesions of the small vessels (arteries and capillaries), particularly in the kidneys Microscopic examination reveals aggregated red cells ("sludged blood") in the small vessels Platelet destruction, intravascular coagulation, and defibrination may occur

Circulation through the microvasculature is critically related to rheological properties of the blood (Merrill, 1969, Wells, 1964) The administration of low molecular weight dextran or albumin has been suggested to improve in vivo the stability of red cells and the other constituents of blood

7 / GRANULOCYTES

Their Production and Role In Defense Against Bacterial Infection

Introduction

In this chapter attention will be devoted to the neutrophilic granulocyte which plays such a central role in the defense against bacterial infection Clinical experience demonstrates the increased susceptibility to infection of individuals in whom the neutrophilic granulocyte count is depressed below 1000/mm ^{3} This is particularly evident following whole body irradiation of man or animals For example, one of the major causes of death in the Japanese exposed to ionizing radiation from the nuclear bombs at Hiroshima and Nagasaki was bacterial infection that followed aplasia of the bone marrow and marked depression in the granulocyte counts in the peripheral blood (Bond et al , 1965) There are also valuable drugs in the physician's armamentarium that unfortunately produce bone marrow aplasia with granulocytopenia The role of the neutrophil in defense against infection has long been known Its role in inflammation and defense was enunciated by Metchnikoff (1887) and emphasized by the ubiquitous presence of granulocytes in pus from all sources

The nucleus of an adult granulocyte is composed of two or more lobes connected by strands of chromatin The cytoplasm contains granules that stain in a characteristic fashion They are of three types *1)* The *eosinophils* constitute only about 2 to 4% of the total white cell count normally, their granules are oval, coarse, and stain bright red with acid dyes such as eosin These leukocytes increase in number during allergies, parasitic infections and idiopathic diseases *2)* The basophils are present to the extent of only 0 15% or less, their granules are large and coarse and stain deeply with basic dyes, e g , methylene blue Their significance is not clear, although they appear to be related to heparin and histamine metabolism Eosinophils and basophils will not be discussed further here *3)* The *neutrophils* are by far the most numerous leukocyte in the blood of human beings. Blood leukocytes of a normal individual are 65 to 70% neutrophils (the remaining being mononuclear leukocytes, that will be discussed in Chapter 9) The neutrophil granules are fine and stain a violet tint with neutral dyes The neutrophils are actively ameboid, are capable of locomotion, and have very active phagocytosis They measure about 10 to 14 μm in diameter Their nuclei show a variable number of lobes The number of lobes is called the Arneth count—the greater the number supposedly the older the cell This is an archaic concept and not useful

GRANULOCYTE STRUCTURE

Structurally the nuclei of all mature peripheral blood granulocytes are similar by light and electron microscopy Nuclear pores are seen, but they are not as prevalent as in the more mature of those cells that are still capable of mitotic division There is no clearcut evidence of transport between the cytoplasm and the nucleus There is no well-defined nucleolus Since in other cells the nucleolus is the origin of much of the cytoplasmic ribonucleic acid (RNA), it is probable that in neutrophils most major biosynthetic processes have ceased (Zucker-Franklin, 1968) Similarly, the Golgi body and centrioles are elementary and found with great difficulty

Granules are membrane-bounded structures containing hydrolytic and oxidative enzymes characteristic of lysosomes (Cohn and Hirsch, 1960) The neutrophil is a mature, motile cell laden with potent enzymes awaiting the opportunity to perform its one obvious function of phagocytosis, a process discovered by Metchnikoff (1887) During phagocytosis an orderly sequence of events takes place The granulocyte surrounds the particle to be phagocytized with pseudopods which fuse, engulfing the particle into a vacuole limited by cell membrane Thus, the particle still occupies extracellular space although inside of the cytoplasm The neutrophilic granules (lysosomes) fuse to the membrane and inject their content of hydro-

lytic and oxidative enzymes into the phagocytic vacuole This process is paralleled by degranulation (Zucker-Franklin, 1968)

During the process of phagocytosis there is a burst of both aerobic and anerobic metabolism, although glycolysis is believed to provide the energy necessary for this phagocytic activity (Karnovsky 1962) The hexosemonophosphate shunt is rapidly activated Hydrogen peroxide is generated Oxygenases that use molecular oxygen are present within the leukocyte and may be important in bacterial killing After the phagocytosis and burst of enzymatic and respiratory activity, the hydrolytic enzymes are activated by a fall in the pH

Phagocytosis of a single bacterium causes a limited amount of degranulation Phagocytosis of 50 bacteria may produce complete degranulation and death of the phagocytic granulocyte that releases the lysosomal enzyme and histamine, and contributes to the various extracellular mediators of inflammation (Cline, 1970) A large number of granulocytes disappear from the blood daily and are destroyed without producing any detectable physiological or harmful effects Actually, the average number of granulocytes destroyed per day in standard man (70 kg) is about $1\,14 \times 10^{11}$ This represents 1600/mg /day or 66/mg /hr averaged out through the body Krogh (1922) estimates the filtration bed of the capillaries to be 6300 m^{-2} If granulocytes leave through the capillary bed and die in the tissues this represents 18/day/mm^{-2} of capillary bed The capillaries of the alveolae of the lungs are estimated to have a surface area of 140 m^{-2}. If granulocytes normally proceed through capillaries in proportion to the area of capillary bed about 4 5% of these granulocytes stray into the alveoli The lungs, however, are only 1 4% of body weight, and the alveoli represent a small fraction of lung weight Hence the alveoli are richly endowed normally with a large supply of granulocytes for a first line of defense against airborne bacteria

Under some circumstances it is believed that the powerful enzymes of the neutrophil may be released and produce undesirable or severe painful pathological episodes For example in gout, urate crystals are deposited in joints and joint capsules producing mechanical irritation Neutrophils migrate into the area and phagocytize the crystals In the absence of an enzyme, uricase, the neutrophil is injured, empties its enzymes and humoral inducers of inflammation into and around the joints and produces a violent painful inflammatory response (Malawista 1965) It is also believed that chemotactic influences e g antigen-antibody reactions may attract neutrophils into tissues (joints in rheumatoid arthritis, arteries in periarteritis nodosa kidney in glomerulonephritis, etc) resulting in death of the neutrophil Subsequent release of their enzymes (including collagenases and elastases) may be the cause of the necrotizing and severe inflammatory lesions that typify these diseases (Janoff, 1960)

The tissue release of these enzymes nevertheless serves a useful purpose at times Neutrophils accumulate in areas of tissue destruction following burns and trauma Through the release of collagenases and other enzymes they accomplish a debridement of necrotic tissues

CHEMOTAXIS OF LEUKOCYTES

Chemotaxis is a term that has been used to describe the directional locomotion of cells towards chemical substances or particles in their environment The chemotaxis of phagocytic cells can be looked upon as a phenomenon of recognizing harmful foreign material such as bacteria or their products The ensuing directional migration of phagocytes towards the inducing agent probably serves as a first stage in the response of phagocytic cells to bacteria When in contact with the material, phagocytosis occurs There are many techniques for the measurement of chemotaxis in vitro and in vivo (Sorkin et al , 1970) Many substances are not chemotactic per se but exert their effect by inducing formation of chemotactic mediators in fresh serum (Keller and Sorkin, 1965) It is believed that antigen antibody complexes and several fractions of complement interact to result in chemotaxis Cationic polypeptides from granules of polymorphonuclear cells increase the vascular permeability, cell adhesion to the vascular walls and infiltration of leukocytes into an inflamed area (Golub and Spitznagel, 1966) One can imagine that the initiation of the inflammatory response by a few granulocytes results in a local killing of a few granulocytes and release of material from their granules, which in turn serves as a multiplication factor comparable to a chain reaction that attracts more

cells into the inflammatory area Chemotactic factors are specific for neutrophil and macrophages, respectively

THE INFLAMMATORY RESPONSE

For centuries medical students have been taught the cardinal signs of inflammation—rubor, tumor, calore, and dolore (Celsus) Inflammation may progress to loss of function Histologically, one sees necrosis of cells, vasodilation locally, edema and accumulation of leukocytes The inflammatory reaction is one of the mechanisms of host defense against attack by bacterial and other agents As alluded to earlier, however, the inflammatory reaction may develop under other circumstances and become harmful to the host The inflammatory reaction is induced not only by cells but also by humoral agents released from cells, and produced by the leakage of specific plasma proteins into the area Classically the cellular events consist of margination and sticking of leukocytes to the vessels, their transit through the vessel walls and migration into the inflammatory area In general pathology, the acute inflammatory process commences with the terminal arterioles, capillaries, and venules The circulation at first is characterized by vasodilation with increased capillary permeability and the leakage of plasma proteins *Histamine* release probably plays an early transient role in the inflammatory process by inducing increased capillary permeability In a later stage, during which leukocytes are infiltrating into the tissues, the blood flow is slowed and hemorrhage may occur *Serotonin* may also be released from platelets during their breakdown within tissue, and also could influence blood flow and leakage of plasma from the capillaries by a combination of arteriolar dilation and venular constriction (Majno and Palade, 1961)

An exceptionally important humoral system in the inflammatory response is the production of *bradykinin*, a linear nonapeptide with a molecular weight of 1060 (Rocha e Silva, et al, 1949, Elliott et al, 1961, Pierce and Webster, 1961) Bradykinin is accepted as the most powerful pain producing agent known (Armstrong et al, 1957) It is exceptionally potent in producing arterial and venular dilatation by direct action on smooth muscle It thus participates in the inflammatory response in a manner similar to histamine and serotonin but is much more potent All three induce separation of intercellular junctions of the venular endothelium, causing endothelial cells to contract and swell into the lumen of the vessel and pulling endothelial cells away from each other and exposing subendothelial tissue (Majno et al, 1969) Bradykinin is a prototype of several low molecular weight polypeptides (*kinins*) which are released from an inactive plasma αglobulin substrate called kininogen by a number of enzymes that exist in the plasm and tissues called kallikreins The kallikreins in themselves must be activated in vivo al though this mechanism is not well-known The intricate interrelations of these systems have been analyzed (Donaldson, 1970, Miller and Melmon, 1970) The activation of the kalli krein-kinin-kininase system is related to a key blood coagulation factor, the Hageman factor (Margolis, 1958, Webster and Ratnoff, 1961) This is shown schematically in Figure 4 24

Granulocytopoiesis

THE BONE MARROW PRODUCTION OF GRANULOCYTES

Approaches to the study of cell proliferation in the bone marrow were described in Chap ters 2 and 3 The characteristics of granulo poiesis as a cell renewal system are shown in Table 4 12 The stem cell pool has been en dowed by the fountain of youth with immortality, and continually replenishes itself and provided cells for all differentiated hemopoietic cell lines by the action of feedback signals that starts the stem cell down an irreversible pathway In contract to the infinate capability of stem cells to divide upon need, the differentiated cell lines have the capability of dividing only a finite number of times before some cytoplasmic nuclear interaction causes them to

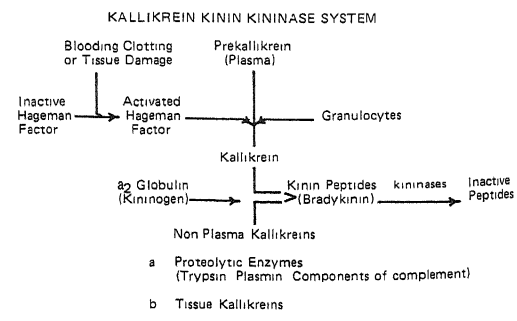


Fig 4 24 Kallikrein-kininase system (From Miller and Melmon 1970)

Table 4 12
Scheme of a cell renewal system and structure of granulopoiesis in parallel indicating function and capabilities

Cellular Pool	Granulopoiesis	Characteristics	Mitotic Capability
Stem cell	Stem cell		Self-perpetuating
Multiplicative	Myeloblast (M_1)	Amplification of input	Finite number of
	Promyelocyte (M_2)	and maturation	mitoses allowed
	Myelocyte ($M_3 + M_4$)		
Maturation	Metamyelocyte (M_5)	Maturation	None
	Band (M_6)		
	Segmented (M_7)		
Functional	Segmented in blood (M_7)	Function fight infection	None
Cell death	Random loss and pyknotic segmented granulocytes	Random or senescent death	

From Cronkite and Vincent, 1969

cease cycling Two important transitions in the flow of cells from the stem cell to the blood granulocyte are shown The first is the invisible but definite flow from the stem cell to the granulopoietic compartment Second, the myelocyte-metamyelocyte transition represents the flow from the multiplicative compartment into the first part of the nondividing maturation compartment This transition is clearly defined by the movement of labeled cells, from the proliferative compartment into the first stage of the nonproliferative compartment The statistical variability of this rate has not been established A single in vivo observation in man was 3 3%/hr (Cronkite et al , 1960), probably faster than normal because of infection The third transition is the emergence of neutrophils into the blood, which can be measured by the radioautographic appearance of labeled granulocytes in the blood The transitions from myeloblast to the promyelocyte and thence to the myelocyte cannot be observed directly, since pulse labeling with tritiated thymidine (^{3}HTdR) simultaneously tags each of these stages The behavior of the stem cell pool and its regulation were discussed in Chapter 2 Recent observations suggest that granulopoiesis at stem cell level is controlled in a manner similar to erythropoiesis with a committed and pluripotent stem cell, the former in cycle and the latter primarily in G_0 (Rickard et al , 1971) The fine details on granulocytopoiesis have been reviewed (Cronkite and Vincent, 1969)

The observations required for a quantitative discussion of granulocytopoiesis are *1*) the proportional distribution of cells within the bone marrow obtained by a marrow differential count of sufficient cells to attain reproducibility in defining the ratio among the cytological compartments, *2*) the absolute cellularity of bone marrow providing the number of dividing and nondividing granulocytic cells of the marrow per kilogram, data essential for computation of absolute production rates in the marrow (Donohue et al , 1958), *3*) the mitotic and/or ^{3}H thymidine labeling index with the respective mitotic and deoxyribonucleic acid (DNA) synthesis times (Strychkmans et al , 1966)

With these data one can estimate the fluxes into, the birth rate within, and the fluxes out of sequential cell compartments relative to the number of the most immature identifiable cell, the myeloblast as described in Chapter 2 In Table 4 13 the proportional distribution of cells in granulocytopoiesis within the bone marrow is tabulated along with the in vitro labeling index, the number in synthesis (N_s), the fluxes into and out of each compartment and the birth rate (K_b) within the compartments In earlier studies the smaller myelocyte M_4 was considered to be a more mature myelocyte than M_3 Present thinking places this cell as primarily a G_1 myelocyte (Cronkite and Vincent, 1969) Accordingly, in the estimation of the relative production M_3 and M_4 are pooled Since the stem cell cannot be identified its flux is omitted, hence, the relative production rates are underestimates Later the stem cell influx will be estimated indirectly The number of cells appearing in M_5 and those exiting M_4 are close There may be a small loss of cells in the production line particularly when the stem cell influx is considered In fact it has been postulated that granulocyte pro-

Table 4 13
Distribution of granulocyte precursors in normal marrow[a]

	M_1	M_2	$M_3 + M_4$	M_5	M_6	M_7
N[b]	100	283	4720	3784	6227	8964
I_L	0 153	0 585	0 22	0	0	0
N_S	15 3	165	1023	0	0	0
K_{in}	?	1 3	15 1	98[c]	98	98
K_B	1 3	13 8	85 2	0	0	0
K_{out}	1 3	15 1	100 3	98	98	98

From Cronkite and Vincent, 1969

[a] Average of five normal marrows after 30 min incubation in vitro with ^{3}HTdR M_1, myeloblast, M_2, promyelocyte, M_3, large myelocyte, M_4, small myelocyte, M_5, metamyelocyte, M_6, band, M_7, segmented granulocyte

[b] Number in each compartment relative to $M_1 = 100$

[c] 2 5/hr the turnover rate required to equal the absolute granulocyte turnover rate in the blood

duction may in part be regulated by a controlled death function in such a way that turning off the death function-increases the output This notion suggests that cells die at the myelocyte level unless needed (Patt and Maloney, 1964)

The relationship among flux out (K_{out}), flux through DNA synthesis (K_s) and stem cell flux into (K_{in}) in a system in which there are multiple mitoses can be expressed thus

$$K_{out} = 2^n K_s = 2^n K_{in} \quad (1)$$

$$K_{out} = [\Sigma (2^{n-1} + 2^{n-2} \quad + 2^0] + K_{in} \quad (2)$$

The first term is equal to

$$\Sigma N_s/T_s = K_B \quad (3)$$

and

$$K_{out} = K_B + K_{in} \quad (4)$$

where

T_s = DNA synthesis time

n = number of mitoses

It follows that K_{in} is also equal to the rate of entry of cells into, through, and out of any phase of the generative cycle. If not, then cells will pile up in a phase, and a steady state will not exist Aberrations in which cells pile up are characteristic of some abnormalities

Equation (3) allows calculation of production rate without knowledge of the number of mitoses because it requires only labeling of cells in DNA synthesis by the specific precursor, tritiated thymidine, and knowledge of DNA synthesis time However, as will be shown later, one can estimate the average number of mitoses utilizing the preceding equations

The preceding concept and the data in Table 4 13 for the ratio of cells in DNA synthesis for myeloblast, promyelocytes, and myelocytes of 85 190 1100, or 1 2 2 13 indicates that this ratio is best fitted by assigning approximately one mitosis to myeloblast and promyelocyte and two sequential mitoses in the myelocyte If each cell were to divide accordingly, there would be an amplification factor of 16 from the stem cell However, there is no direct evidence that each cell committed to granulopoiesis actually divides four times before maturation and entering the first compartment of the nonproliferating pool M_5, the metamyelocyte

The blood granulocyte turnover rate (GTR) is $6\,79 \times 10^7$/kg /hr If there is no loss between the myelocyte and the peripheral blood granulocyte the marrow must make at least $6\,79 \times 10^7$ granulocytes/kg /hr The marrow myelocyte mass is 260×10^7/kg (Donohue et al, 1958) From Table 4 13 it is seen that 0 22 of the myelocytes are in DNA synthesis, hence, the

$$K_B = \Sigma N_s/T_s = (260 \times 10^7 \times 0\,22)/12$$

$$= 4\,76 \times 10^7/\text{kg /hr}$$

The flux into the myelocyte pool must be

$$K_{in} = K_{out} - K_B = 2\,03 \times 10^7/\text{kg /hr}$$

Rearranging equation (1) and substituting,

$$n = \frac{\log_e (K_{out}/K_{in})}{\log_e 2} = \frac{\log_e (6\,79/2\,03)}{\log_e 2} = 1\,74$$

the average number of mitoses at the myelocyte level as derived from the values presently available for GTR The average transit time through the myelocyte compartment will be 91 hr Cronkite and Vincent (1969) deduced that the generation time of myelocyte is about 52 hr Of this the DNA synthesis time is 12 hr, and G_2 + M is 3 hr, leaving 37 hr for G_1 The ratio of DNA synthesis time to generation time is 0 23, close to the observed labeling index of 0 22 given in Table 4 13 These are current estimates not rigorously proved by human data and subject to revision as more information becomes available The examples are used to illustrate the principles involved

In producing 6 79 × 10^7 granulocytes/kg /hr (assuming no loss between the metamyelocyte and the emergence of granulocytes into the peripheral blood and using the absolute marrow cellularity of 270 × 10^7/kg for M_5, 360 × 10^7/kg for M_6 and 250 × 10^7/kg for M_7,) the respective turnover rates must be 2 5%/hr, 1 9%/hr, and 2 7%/hr and the respective turnover times are 40, 53, and 37 hr This is illustrated in Figure 4 25

A major discrepancy exists between the relative data on distribution of nondividing granulocytes in the marrow and the absolute number used in the calculations of the preceding paragraph and shown in Figure 4 25 When similar computations are made on a basis of relative numbers in Table 4 13, the following times are obtained 40, 66, and 94 hr, respectively for M_5, M_6, and M_7 with turnover rates of 2 5%, 1 5 and 1 06%/hr There is no obvious explanation to date for this discrepancy, except that one or another observation is in error

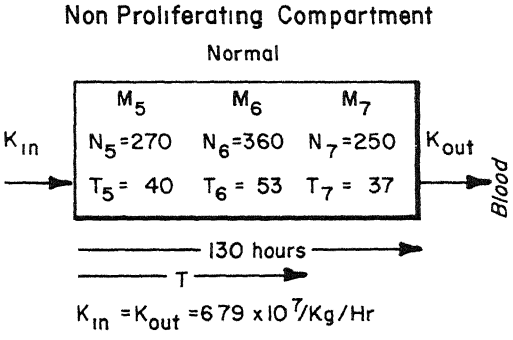


Fig 4 25 Compartment size (N) and transit time (T) for the nonproliferating granulopoietic compartment based on absolute cellularity in cell per Kilogram body weight

The number of segmented granulocytes stored in the marrow is about four times the number in the blood The total number of stored granulocytes in the entire nonproliferating compartment (M_5 + M_6 + M_7) is about 12 5 times the number in the blood Upon demand a physiological mechanism exists which releases the mature granulocytes quickly into the blood and accelerates the maturation of the rest of the storage compartment, with a resulting shortening of the emergence time of cells labeled at the myelocyte level (Fliedner et al 1964b) The chemical factor responsible for the rapid release has been termed the *leukocytosis inducing factor* (Gordon et al, 1960a) A sustained granulocyte turnover rate requires accelerated cell production and release (Lapin et al, 1969)

THE BLOOD GRANULOCYTE

Knowledge of the behavior of the blood granulocyte has accrued primarily from the use of diisopropylflurophosphate ($DF^{32}P$) labeling of granulocytes followed by autotransfusion of the labeled cells and observation of their dilution and disappearance rate (Cartwright et al, 1964) The flow of ^{3}HTdR labeled cells from the marrow into the blood (Fliedner et al, 1964a,b, Cronkite and Vincent, 1969) has also given considerable insight into the behavior of the blood granulocyte The rapid granulocytosis produced by exercise or adrenalin has long been thought to be due to the rapid mobilization of a pool of readily available mature granulocytes in the blood The "margination" of granulocytes in areas of the circulation where the blood flow is relatively slow is an old observation (Vejlens, 1938) From the $DF^{32}P$ labeling of an individual's granulocytes followed by autotransfusion, it has been directly shown by dilution that the blood contains a circulating granulocyte pool (CGP) and a marginal granulocyte pool (MGP), and that the sum of these equals the total blood granulocyte pool (TBGP) The ratio of the CGP to the TBGP is 0 44 with a range of 0 16 to 0 99 An equilibrium between the CGP and MGP is attained rapidly after introduction of labeled granulocytes into the circulation. The disappearance of autotransfused $DF^{32}P$ labeled cells is random with a half-time ($T_{1/2}$) of 6 7 hr (Athens et al, 1961). The notion that the loss of granulocytes of the blood is random is di-

Table 4 14
The total blood granulocyte pool (TBGP) circulating granulocyte pool (CGP) marginal granulocyte pool (MGP) half time disappearance ($T\frac{1}{2}$) and granulocyte turnover rate (GTR) in normal subjects

Determination	No of subjects	Mean	95% limits
TBGP $\times 10^7$ cells/kg	100	70	14–160
CGP $\times 10^7$ cells/kg	109	31	11–46
MGP $\times 10^7$ cells/kg	109	39	0–85
$T\frac{1}{2}$ in hours	56	6 7	4–10
GTR $\times 10^7$ cells/kg /day	56	163	50–340

From Cartwright et al , 1964

rectly supported by the observation that ^{3}HTdR labeled granulocytes appear in the blood and extravascular sites such as saliva simultaneously (Fliedner et al , 1964b) Blood granulocytes also die by a process of senescence This was demonstrated by the appearance of labeled pyknotic (dying) forms seen in the peripheral blood 30 hr after labeled granulocytes first appeared in the blood Thus, the random loss of granulocytes from the blood is truncated by senescent loss about 30 hr after the entrance of the cells into the blood (Fliedner et al , 1964b) In Table 4 14 the values on measurement of the various compartments of the blood granulocyte pool are tabulated The fractional turnover rate per hour (λ) for a compartment from which there is random loss is $\log_e - T\frac{1}{2}$ The reciprocal of λ is the average time spent in the compartment The granulocyte turnover rate (GTR) is therefore,

$$GTR = TBGP \times \lambda = 6\,79 \times 10^7/kg\ /hr$$

$$\lambda = 0\,103/hr$$

$$1/\lambda = 9\,7\ hr$$

The GTR is $6\,7 \times 10^7$/kg /hr or 163×10^7/kg /day

It has always been tacitly assumed that the instantaneous marrow production rate (MPR) equals the marrow release rate (MRP) equals the granulocyte turnover rate (GTR) in the blood The notion of such a steady state where instantaneous production and turnover rates are equal is incompatible with the observation of a long term oscillation of the granulocyte count in the blood of normal men (Morley, 1966). From peak to peak the granulocyte oscillation in the peripheral blood varies from 14 to 23 days This cycling is considered as being compatible with a system that has a negative feedback circuit acting on the marrow stem cell (self perpetuating level) with a time delay corresponding to the time from induction of the stem cell into granulopoiesis until it appears in the blood as a granulocyte The time delay would approximate half of the cycle time Intuitively one might expect a normal cycling of the blood granulocyte concentration since many tissues are in a state of flux, the size varying with changes in rate of production on one hand or in the rates of disappearance on the other hand, and regulated by need and use With the reasonable establishment of a granulocyte cycle in the blood, steady-state calculations of granulocyte kinetics should be based on the average production and granulocyte turnover rate through an entire cycle rather than at any single point in time MPR will equal GTR at some time later when the cells produced at time (t_1) are being destroyed in the blood The periodic state is shown in Figure 4 26 Although the actual curve of the variation of the granulocyte concentration with time (solid line) does not conform to a sine wave it is shown thus for simplicity The reciprocal of this curve, the dotted line, represents the strength of the signal that drives stem cells down the granulopoietic pathway This signal presumably initiates a complex sequence of

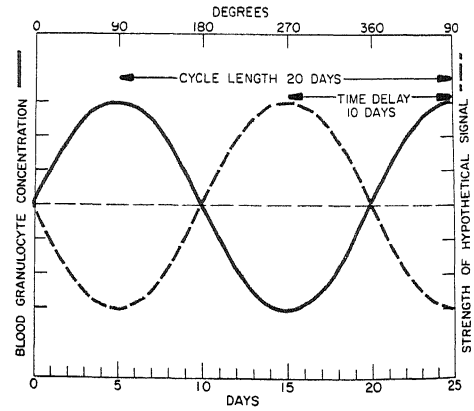


Fig 4 26 Diagram of concept of cyclic variations in the blood granulocyte concentration based on the observations of Morley Granulocyte concentration is shown as a solid line and the strength of a hypothetical signal stimulating granulopoiesis is shown as a broken line Sine waves have been used in this model although in fact the shape of the curves is probably more complex than this The cycle length is 20 days and the time delay from maximum signal to maximum response is 10 days

biochemical events that constitute the process of differentiation into a cell type characterized by the numerous specific proteins enzymes granules, and organelles as described in the earlier part of this chapter In keeping with contemporary knowledge on transfer of biochemical codes, the signal is almost certainly chemical and initiates the derepression of transcription sites for m-RNA coded for the enzymes and proteins that are characteristic of the cell line Since many specific proteins and enzymes are going to be synthesized it is conceivable that the chemical signal derepresses the transcription sites for a single protein, and that the subsequent cascade of derepression of transcription sites for different proteins is related to repression of other sites by successive protein products These, in turn, after attaining a specific concentration may suppress their own transcription sites by product ' inhibition, analogous to that suggested for mean corpuscular hemoglobin concentration in turning off cell cycling in erythropoiesis (Chap 3)

Several factors influence the TBGP, CGP, MGP, and GTR Adrenal corticosteroids increase the total granulocyte count apparently by a combination of increase in MRR and a decreased efflux from the blood (Bishop et al 1968) Bacterial endotoxin produces a granulocytosis preceded by a fleeting granulocytopenia with shift in cells from the CGP to MGP This is followed by an absolute increase in TBGP with a decrease in granulocytic specific activity of the labeled cells, indicating an increased MRR of unlabeled cells (Athens et al , 1961) The TBGP is increased in infection where the half-time in the blood is usually prolonged, and GTR is always increased This increased GTR demands an increase in MRR and MPR as described earlier

When there is a sustained granulocytosis the mean age of the granulocyte in the blood measured from the last myelocyte mitosis is younger Probably these cells have not reached the degree of effeteness described in the first section on the granulocyte structure, and thus, may have a longer life that perhaps explains the longer half-time of granulocytes in the blood during infection

The reserve granulocytes available to be drawn on in time of need are the mature granulocytes in the marrow, the number of which is about four times the number present in the peripheral blood The total nonproliferating pool of marrow cells is about 12 5 times the number present in the blood Cells leave the marrow more or less by a "pipeline ' process the older cells being released first (Maloney and Patt, 1968) The time needed for a cell labeled with tritiated thymidine as a myelocyte to appear in the peripheral blood is defined as the emergence time it varies between 96 and 144 hr normally It can be shortened to 48 hr under the stress of bacterial infection (Fliedner et al 1964a)

FACTORS RELATED TO REGULATION OF GRANULOPOIESIS

In Figure 4 27 an idealized scheme of granulopoiesis is presented The ordinate shows the flux of cells as 10^7/kg /hr and the abscissa the time from entry of a stem cell into granulopoiesis From the stem cell to the metamyelocyte the transit time is 138 hr During the normal periodic state the stem cell influx will vary as described earlier, and will give rise about 288 hr later to a rising or falling input into the peripheral blood, which in turn determines the concentration of granulocytes in the blood The rigid scheme shows a 16-fold amplification of the stem cell input if each cell that enters granulopoiesis shows four mitoses and divides four times Some cells may divide fewer times as suggested in earlier discussion As mentioned earlier, the myelocyte is believed to have a G_1 period of 37 hr A theoretical mechanism of increasing production rate in the same time interval from stem cell to metamyelocyte would be a shortening of G_1 Sufficient time would be provided for an additional mitosis and thus, a doubling of the production rate within the 138 hr transit from the stem cell to the metamyelocyte Whereas, the transit from the last myelocyte mitosis (emergence time) normally ranges from 96 to 144 hr , during the stress of infection the emergence time can be decreased to as little as 48 hr (Cronkite et al , 1960) Thus, with a shortening of the generation time, an additional mitosis and a shortened transit through the nonproliferating pool an increased input into the blood could be experienced within a few days These three mechanisms constitute an attractive hypothesis but one not yet experimentally verified

If the entire increase in production were to

come from an increased input from the stem cell level without changes in generation time, it would take about 12 days to be seen in the peripheral blood Inasmuch as sustained granulocytosis is seen much faster than this, an increased stem cell input does not appear to be the sole mechanism of increasing granulocyte production Prolonged increased GTR requiring more than twice the normal production will demand an expansion in the mass of the granulopoietic marrow

In Figure 4 28 the notions discussed are in-

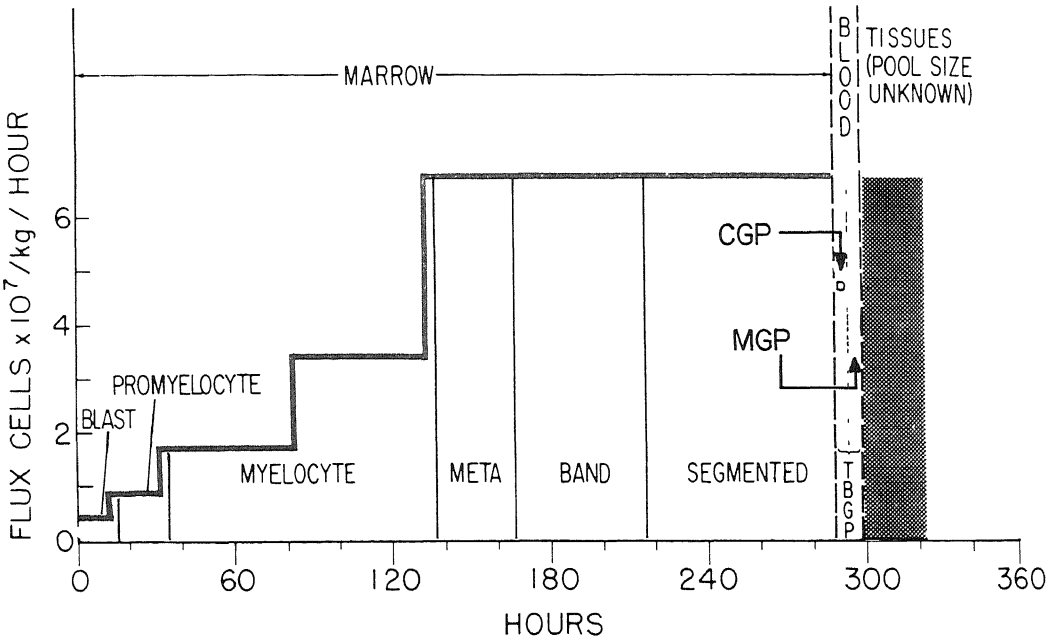


Fig 4 27 Sequential flow of cells through granulopoiesis The ordinate shows the flux through each compartment relative to the GTR and the abcissa shows the time in each compartment The area of each of the nondividing compartments gives the number of cells in that compartment In the case of the proliferating compartments large and small myelocytes have been combined The stepwise increase in cell numbers through the dividing compartments represents serial divisions The size of TBGP is in scale but the relationship between CGP and MGP does not necessarily imply flow through MGP to the tissues The size and existence of the tissue pool are unknown (From Cronkite and Vincent 1969)

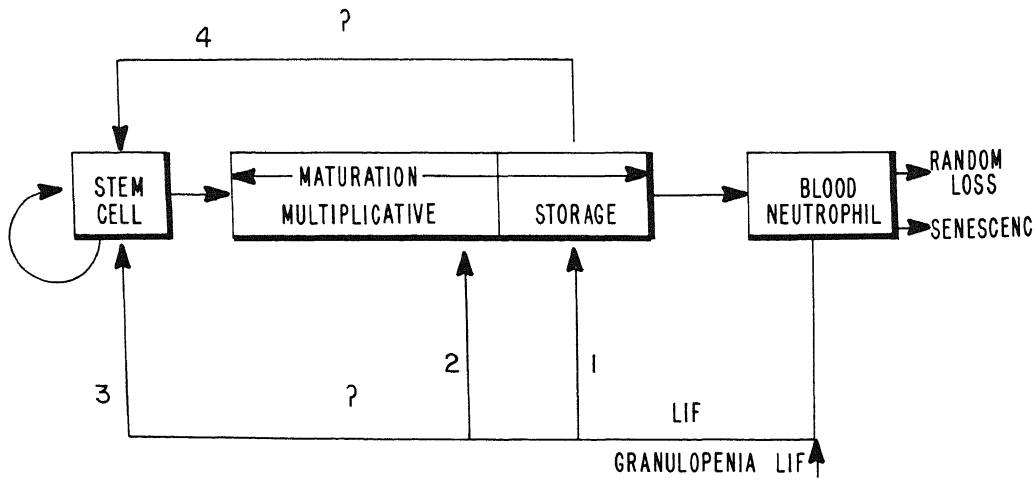


Fig. 4.28 Concept of regulation of granulopoiesis by several possible feedback loops *1)* LIF releases stored marrow granulocytes *2)* Myelocyte generation time decreases with more mitoses *3)* A peripheral signal differentiates more stem cells and triggers more cycling to restore size *4)* Possible intramedullary signal from storage to stem cell pool induces differentiation and cycling

corporated into a general operational concept of regulation of normal granulopoiesis and its response to stress Granulocytes are being lost from the blood by a random and senescent process As the granulocyte count falls from any reason the leukocytosis inducing factor (LIF) increases in concentration (Lapin et al 1969) The LIF through a fast feedback loop causes the marrow to release the storage granulocytes quickly Then either due to the action of LIF or other factors, it is believed that the myelocyte generation time decreases with an additional mitosis that doubles the output The transit through storage compartment is shortened Then through a feedback loop to the committed or possibly pluripotent stem cell, there is an increased rate of differentiation followed by increased rate of proliferation to rebuild the stem cell pool size Whether this results from an intramarrow sensing device that determines the diminution in size of the storage compartment or is a direct action of LIF or other peripheral factor on the stem cell pool is obscure

CLINICAL ASPECTS OF GRANULOPOIESIS

Dangers of Chemotherapy For a computer analysis a model for granulopoiesis can be based on the following concepts a periodic cycle of blood granulocyte concentration, regulated by two feedback loops—a fast loop that controls the release of stored granulocytes in the bone marrow (LIF), and a slow loop that controls the entrance of stem cells into the granulopoietic pathway, the time constant observed for lifespan of cells in the peripheral blood, and the transit time through the bone marrow This computer model reproduces the low amplitude oscillations observed in normal males (Morley et al, 1970) In addition, it yielded an important prediction to the effect that mild marrow failure might produce a cyclic neutropenia with a large amplitude This has been experimentally confirmed With a decreasing store of marrow granulocytes the fast feedback loop becomes less capable of modulating the inherent tendency of a system with long time delay to oscillate, and one can therefore detect swings of larger amplitude With further reduction in size of the proliferative compartment of the marrow, the production rate of granulocytes will fall further, and as the size of the marrow granulocyte reserve approaches zero, the extreme oscillations will flatten out into a noncyclical neutropenia of severe and dangerous degree Such has been frequently observed after prolonged heavy dosage of marrow depressant drugs or radiation Clinically the early detection of marrow depression before it is irreversible is of utmost importance in radio or chemotherapy of metastatic disease in procedures for immunosuppression, or in evaluation of possible marrow toxicity of new drugs To safeguard the patient an adequate number of granulocyte counts should be performed at intervals sufficiently frequent to reveal any increase in the amplitude of any granulocyte cycle that might precede the development of a severe noncyclical neutropenia Isolated counts can be misleading if they are made by chance at the peak or the bottom of a cycle

Chronic Granulocytic Leukemia. Chronic granulocytic leukemia is clearly an aberration of granulocytopoiesis There is an increase in the blood concentration of leukocytes at all levels of maturation, and proliferation from the marrow into the peripheral blood with cells' being seeded out in the spleen and elsewhere where they proliferate This disease could be due to an increased input of stem cells that enlarges the entire system and expands the marrow compartment to the point that it 'leaks'' and cells metastasize to the body at large There is also an enzyme defect of granulocytes in chronic granulocytic leukemia with the absence of alkaline phosphatase In many cases, the abnormal Philadelphia chromosome described in Chapter 2 is present

Marrow Suppression by Drugs and Radiation. Following exposure to whole body irradiation there is a decrease in the stem cell input into granulocytopoiesis that is followed by a progressive depletion in the maturation and multiplicative compartment and is finally reflected in the blood as a marked neutropenia If the irradiation has been sufficiently heavy and blood cell production has decreased to near zero levels the individuals are markedly susceptible to infection and may die of this cause before regeneration of the bone marrow can occur (Bond, et al, 1965)

Drugs can also interfere with blood cell production Butazolidin is an example of a drug that occasionally produces a very severe granulopenia, apparently because of interference in the transit of cells from the myelocyte to the metamyelocyte Examination of the bone marrow will show a striking depletion of the

nonproliferating compartment with an active myelopoiesis. Upon withdrawl of the drug, one can observe an orderly flow of cells through the nonproliferative pool and the reappearance of granulocytes in the blood.

Chronic Granulomatous Disease of Childhood. This is a fatal genetically determined disorder effecting male children with an onset of manifestations early in life (Good et al., 1968). The abnormality involves enzymatic deficiencies in the lysosome phagocytic vacuole system that prevent the effective killing of phagocytotized pathogenic bacteria. The fundamental metabolic lesion in this disease is deficient production of hydrogen peroxide during phagacytosis (Baehner et al., 1970).

Congenital Neutropenia. This relatively benign disorder is characterized by an inability of immature myeloid cells to proliferate and mature to normal numbers of granulocytes (Vriedt, et al., 1970). If the production rate is so low that there is no, or little, storage pool in the marrow a noncyclical neutropenia would be predicted on the basis of the discussion outlined above under dangers of chemotherapy.

THROMBOPOIESIS, HEMOSTASIS, COAGULATION, AND THROMBOSIS

Thrombocytes

The platelet is a non-nucleated cell yet, it has a complex structure The surface limiting plasma membrane is 70 to 90 A in thickness it is interpreted to consist of one lipid bimolecular layer and two protein layers Mitochondria are always present in the cell There are electron dense granules (α granules) that are believed to contain platelet factor 3 and enzymes Vacuoles and vesicles are present, and remnants of Golgi bodies are seen Microtubules can be identified and separated, their function is obscure but appears to be related to maintaining the shape of the platelet Possibly the microtubules represent a retractile element of the platelet cytoplasm called *thrombosthenin* There are abnormalities in platelet structure in various hemorrhagic diseases The ultrastructure of the platelet and its relationship to function has been extensively reviewed (Hovig, 1968)

HEMOSTASIS

Hemostasis is the business of platelets This may be defined as the arrest of blood flow from a break in a vessel wall and the prevention of spontaneous bleeding This function probably requires three properties of platelets *1*) an endothelial supporting function of platelets measured by capillary resistance tests, *2*) a capability of forming hemostatic plugs, which can be measured by the bleeding time in human beings or by the enumeration of red blood cells in the lymph in animals with cannulated thoracic ducts, and *3*) the release of a lipoprotein material called *platelet factor 3* that activates plasma prothrombin to thrombin in the coagulation mechanisms, and can be measured by prothrombin consumption time and other in vitro tests Platelet factor 3 will be discussed later under blood coagulation

The platelets occupy a central role in the staunching of blood loss from small and large blood vessels Following injury to a blood vessel the following sequence of events is observed Blood oozes out through the hole or severed vessel the vessel constricts the endothelium everts, and platelets adhere to the collagen in the underlying connective tissue exposed to the flowing blood by the injury Platelets also migrate into spaces which appear between endothelial cells (Ashford and Freiman, 1967 French et al , 1964) Platelets continue to accumulate on top of the first platelets already adhering to the collagen and injured endothelial cells probably as a result of the release of adenosine diphosphate (ADP) from the platelets themselves or possibly from damaged vascular tissue (Gaarder et al , 1961, Born and Cross, 1964) Platelets continue to accumulate and initiate coagulation in the area By a combination of constriction of blood vessels formation of the hemostatic plug of platelets, and fibrin clot the outflow of blood is decreased and usually stopped Retraction of the clot holds the edges of vessels and wound together and so prevents further loss of blood For greater detail on the hemostatic process see the extensive review on blood coagulation (Donaldson, 1970)

The relationship of the platelets to vascular integrity is inferred, from the fact that spontaneous hemorrhage develops when the blood platelet count falls below 20,000/mm ^{3} (Gaydos et al 1962) At 50,000/mm ^{3} or more a reasonable hemostasis is possible during surgical procedures The exact mechanism whereby platelets prevent hemorrhage and maintain capillary strength is not completely understood but has been reviewed extensively (Johnson et al , 1966) Data obtained by utilizing radiosulfate labeled platelets suggests that the platelets directly contribute something to the capillary endothelium (Cronkite et al , 1960) Platelets

have been observed by electron microscopy to enter endothelial cells in thrombopenic states (Johnson et al , 1964)

Freshly separated platelets transfused into the fatally irradiated dog will prevent hemorrhage (Cronkite et al , 1952) Platelet transfusions have been shown to be efficacious in control of thrombopenic bleeding in man (Hirsch and Gardner, 1952)

MEGAKARYOCYTOPOIESIS AND THROMBOPOIESIS

A model for megakaryocytopoiesis is presented in Figure 4 29 From the studies of Whang et al , (1963) it is clear that the common stem cell feeds the megakaryocytic pool upon need There is also good evidence for existence of a committed stem cell pool as described earlier under erythropoiesis and granulocytopoiesis Megakaryocytes have been divided into three stages depending upon their maturity Tritiated thymidine initially labels stage I megakaryocytes The label can be seen to appear successively in stages II and III As the label progresses through the more mature stages, the percentage labeling in stage I increases, this indicates an influx of cells from an earlier, labeled, unknown precursor, possibly from a committed stem cell compartment As the megakaryocytes mature they produce platelets and die For details of megakaryocyte proliferation, kinetics of their maturation and platelet production see Ebbe (1968)

Platelets are formed when the mature, granular megakaryocytes shed their cytoplasm The normal human platelet has an average volume of 5 8 μm^3 New platelets are probably larger than platelets that have aged Platelet counts are remarkably stable over long periods of time, signifying that production and destruction under normal conditions do not vary significantly (Brecher et al , 1953)

Platelet *turnover* rate must be measured indirectly from measurements of platelet *survival* and platelet *count* The most commonly used method for measuring platelet survival is based on a chromium-51 labeling with auto transfusion The normal platelet turnover is in the vicinity of 40,000/mm ^{3}/day (Davey, 1966) There appears to be some random loss of platelets in addition to the senescence that sets an upper limit on lifespan The mean platelet life is in the vicinity of 9 to 11 days in normal individuals Random removal of platelets appears to be in the vicinity of 5 to 15%/day This apparent random loss may be due to elution of the label

Platelets can be counted by electronic methods, but the most practical laboratory and clinical method remains enumeration with phase microscopy (Brecher et al , 1953)

The peripheral blood platelets appear to be distributed between the blood and a splenic pool It has been estimated that about one-third of the total platelet mass is concentrated in the spleen in equilibrium with the remaining two-thirds distributed throughout the rest of the vascular system (Aster, 1966)

The regulation of platelet production has been studied by hypertransfusion of platelets and the transfusion of plasma from markedly thrombocytopenic animals and human beings These and other studies on regulation were critically reviewed by Ebbe (1968) It can be concluded that there is a feedback regulatory system, in which the level of the circulating platelets influences the rate of megakaryocytopoiesis and platelet production Although it is not conclusively proved, there is evidence suggesting that a thrombopoietin exists which

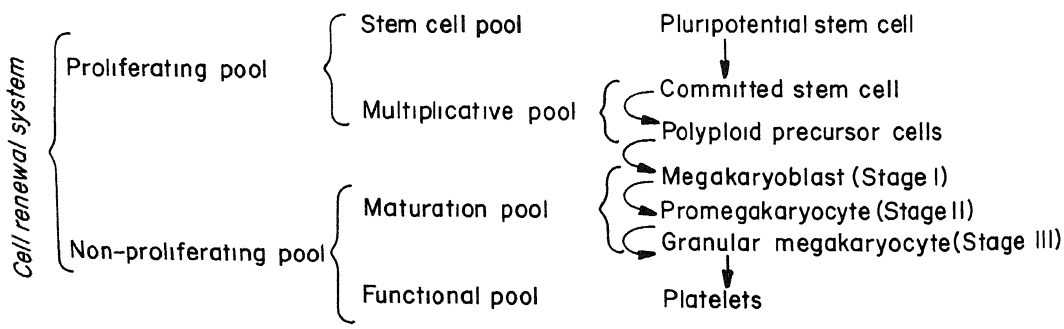


Fig 4 29 Model of megakaryocytopoiesis (From Ebbe)

stimulates platelet production The inhibition of platelet production by increasing concentration of platelets in the blood remains a possibility Because the compartment of recognizable megakaryocytes is a nondividing compartment (Ebbe and Stohlman, 1965) there are two possible mechanisms for increasing the output of platelets 1) each megakaryocyte may make more platelets, or 2) an increased input of cells into the megakaryocytopoietic compartment from the precursor pool may yield more platelet-producing cells It appears that the latter is the mechanism by which increased platelet production is accomplished in acute thrombocytopenia

Coagulation of Blood

GENERAL DESCRIPTION

Clot Formation If blood is collected into a test tube, at the end of a few minutes it will have lost its fluidity and set into a jelly Microscopically this clot is made up of a mesh of very delicate fibrils among which are entangled, as in a net, the red and white cells and many fragmented platelets The fibrils are composed of fibrin that is formed by the alteration and polymerization of fibrinogen molecules If the clot is allowed to stand for an hour or so, it shrinks and expresses from its interstices a clear faint straw-colored fluid This is the *serum*

In this retraction of the clot or *syneresis*, as it is termed, the platelets play an essential role They attach themselves to the fibrin threads which then contract, draw the fibrin compactly together, and thus, squeeze out the serum Clots formed in blood deficient in platelets are soft and friable and easily loosened from the injured vessel The importance of these small particles of protoplasm in hemostasis will be dealt with later

The Clotting Mechanism. No field of physiology is more complex or has been so confused by contradictory results, differences of interpretation, and diversity of opinion as that of blood coagulation Blood coagulation is essentially a series of enzymatic reactions involving a number of plasma proteins Many of these are present only in traces Within an intact healthy circulatory system these enzymes, for the most part, are in an inactive form However, when blood is shed or placed in contact with a "wettable surface," such as glass, a series of autocatalytic enzymatic reactions is initiated In the effort of various investigators to understand the order in which the numerous factors enter into the process of coagulation, the results of an enormous number of 'mixing' experiments have been reported This has resulted in a confusion not only of terminology but also between what is a separate enzyme entity and what is merely a resultant activity To make some order out of chaos an international system of numbering has been introduced for the identification of those factors accepted by a majority of authorities in the field In Table 4 15 the international numbering system for the accepted coagulation factors is shown No numbers have been suggested for the physiological anticoagulants

DEVELOPMENT OF THROMBOPLASTIN ACTIVITY

By the end of the 19th century, it was established that extracts of tissues free of blood accelerate the clotting of whole blood or plasma but have no effect on solutions of *fibrinogen* According to the hypothesis put forward at that time, this tissue extract (called *thromboplastin*), in conjunction with calcium acts by converting *prothrombin* to its active form *thrombin* During the early 1940's it was shown that thromboplastin requires both lipid and protein constituents for full activity Also, at about the same time, work in a number of laboratories indicated that factors other than

Table 4 15
Synonymous terms for clotting factors

International Committee Nomenclature	Synonymous and Descriptive Terms
Factor I	Fibrinogen
Factor II	Prothrombin
Factor III	Thromboplastin
Factor IV	Calcium
Factor V	AC-globulin, proaccelerin labile factor
Factor VII	Proconvertin, SPCA
Factor VIII	Antihemophilic factor
Factor IX	Plasma thromboplastin component (PTC), Christmas factor
Factor X	Stuart-Prower factor
Factor XI	Plasma thromboplastin antecedant (PTA)
Factor XII	Hageman factor, contact factor
Factor XIII	Fibrin stabilizing factor, fibrinase

thromboplastin are involved in the normal conversion of prothrombin to thrombin (Biggs and Macfarlane, 1962)

With the introduction of silicone, a new phase in the study of blood clotting began With proper solutions of silicone, test tubes, syringes, and needles can be covered with a smooth "nonwettable" surface This makes it possible to withdraw blood and prepare plasma for study without causing changes due to surface action Using siliconed glassware it is possible to show that blood itself contains all the factors needed for thromboplastin generation (Biggs and Douglas, 1953), it requires only a change in surface contact to trigger the mechanism Thus, there are two sources of thromboplastin activity, one from tissue extracts sometimes called the *extrinsic source*, and the other from the interactions of a number of trace plasma proteins, calcium ions, and platelets called the *intrinsic source*

The factors that are involved in the formation of intrinsic, or plasma thromboplastic activity are illustrated in Figure 4 30*A* It is now generally accepted that factor XII when activated by surface contact initiates the reaction Then, in the presence of calcium ions, other trace proteins, factors XI, IX, and VIII (possibly in this order), are in turn activated Factor VIII (antihemophilic factor) is present in normal fresh plasma but is absent or inactive in normal serum and rapidly disappears from stored blood, it is relatively stable however, in plasma stored at −20° C, it is not readily absorbed on the usual prothrombin-absorbing reagents such as aluminum hydroxide or tricalcium phosphate Factor IX (Christmas factor) is present in both plasma and serum and is readily absorbed on prothrombin-absorbing agents, it is present in stored blood Deficiencies in factor VIII or factor IX result in a bleeding condition known as parahemophilia, this is an hereditary condition linked with a defect of the X chromosome Factor XI (plasma thromboplastin antecedent) has some of the characteristics of both factors VIII and IX, it is not readily adsorbed on prothrombin-adsorbing reagents but is stable and is found in serum Bleeding defects due to lack of factor XI activity are quite rare In addition to these, and possibly other as yet unidentified proteins, a phospholipid is required for development of optimum thromboplastin activity. During normal coagulation

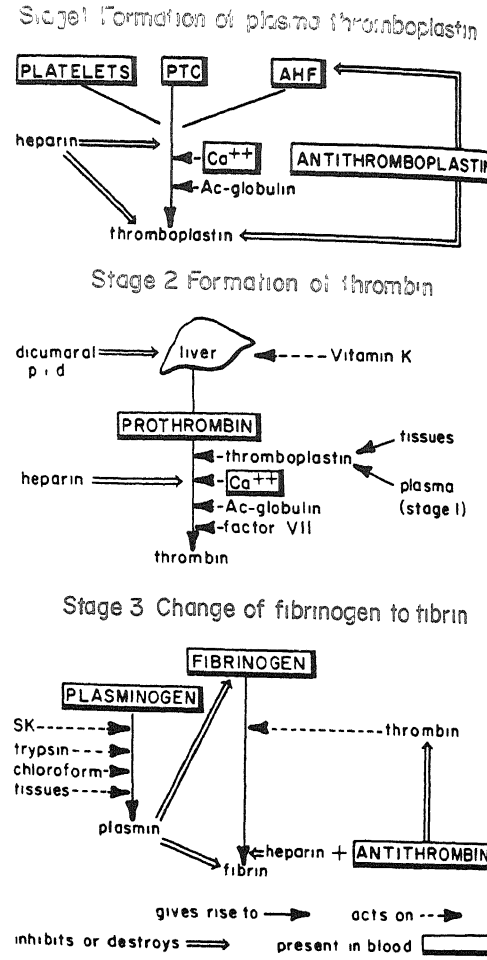


Fig 4 30 Sequence of reactions in clotting mechanism Stage I Formation of plasma thrombo + plastin Stage II Formation of Thrombin Stage III Change of fibrinogen to fibrin

this is supplied through platelet breakdown For in vitro coagulation the platelet phospholipid moiety can be replaced by phospholipid from brain tissue or soy bean (Macfarlane et al , 1964, Owren, 1947)

Agents such as citrate, oxalate, or ethylenediamine-tetra-acetic acid (EDTA) by binding calcium ions will completely inhibit the development of thromboplastin activity Heparin, which is not present in normal blood in measurable amounts, is probably of no importance in preventing normal blood coagulation When added as a therapeutic agent, however, it acts at a number of sites along the chain of reaction There is good evidence that its initial action is to prevent the development of thromboplastin activity It probably does this by com-

bining with the reactive sites on the tract proteins and by preventing platelet disintegration and release of phospholipids. Heparin is a mucopolysaccharide prepared from beef lung or beef and hog intestinal mucosa. A number of similar mucopolysaccharides have been synthesized (Howell and Holt 1935, Best et al. 1938, Jorpes 1949, Monkhouse and Clarke 1957).

ACTIVATION OF PROTHROMBIN

Recent studies have favored the hypothesis that both the thromboplastin formed from the interaction of plasma proteins and the thromboplastin in tissue extracts need an additional factor for full activity. This is factor X (Stuart-Prower factor). In addition to factor X, tissue thromboplastin requires factor VII (convertin). The exact order in which these proteins enter into the reaction is not yet established, but results obtained by Macfarlane and his colleagues (1964) support the order suggested in Figure 4-30*B*. Calcium ions are also required for prothrombin conversion, and factor V (AC-globulin) increases the rate of the conversion. It is known that heparin when added to a physiological system will inhibit the conversion of prothrombin, but much, if not all, of this inhibitory activity may be through the action via thromboplastin. In common with most other plasma proteins, prothrombin, factor X and factor VII are formed in the liver. The ability of the liver to maintain normal blood levels of these protein fractions, however, depends on the presence in the liver of sufficient amounts of vitamin K. Dicumarol, and other similarly acting drugs, antagonize vitamin K and depress the rate of production of prothrombin, factor VII, and factor X (Quick, 1945, Link, 1944).

CONVERSION OF FIBRINOGEN TO FIBRIN

Once thrombin is formed it acts enzymatically to convert the soluble fibrinogen into insoluble fibrin. It has been established that thrombin is a proteolytic enzyme having the same sequence of amino acids at its active site as digestive enzymes such as pepsin and trypsin. Most proteolytic enzymes break down fibrinogen completely, but thrombin splits off only a few terminal peptides. Two of these known as fibrin peptides A and B are the result of selective action of thrombin at terminal arginine-glycine bonds. The splitting off of the peptides causes a change in charge at the ends of the molecules of fibrinogen and allows them to polymerize end-to-end in long insoluble strands (Laki, 1962). When purified fibrinogen is clotted by purified thrombin, the clot is found to be soluble in urea, but the clot formed in plasma is not soluble in urea. A factor has been isolated from human and animal plasma which will render the clot formed from purified fibrinogen insoluble in urea. This factor has been named fibrin stabilizing factor or factor XIII. It greatly facilitates the healing of wounds, yet is missing from the blood in some persons.

Many authorities in the field of coagulation adhere to the view that fibrin is being continually formed and destroyed and that the integrity of the endothelium depends on the presence of a physiological mechanism for the constant removal of fibrin. Such a mechanism does indeed exist. It depends on the presence in blood of an inactive proteolytic precursor called *plasminogen* which has been identified as a β-globulin with a molecular weight in the region of 150 000. During clotting or purification of fibrinogen it is associated closely with the fibrinogen molecule. It is extremely difficult to prepare fibrinogen free of plasminogen and a normal clot usually contains sufficient amounts of plasminogen to insure clot dissolution if circumstances permit its activation. Activators that convert plasminogen to plasmin are of two different types. Activators from tissues and urine act directly on both human and bovine plasminogen. Activators such as streptokinase (prepared from streptococcal cultures) and lysokinases from tissues and blood must act first on a globulin fraction called proactivator. This proactivator is present in large amounts in human plasma, but is virtually absent from the blood of cattle and other animal species. There is usually an excess of plasmin inhibitor in normal blood. However, when large amounts of plasminogen are activated (this can occur during surgery, childbirth, or prostatic cancer), fibrinogen as well as fibrin is destroyed by the enzyme and serious hemorrhage can result. Small increases in the fibrinolytic activity of plasma have been reported following exercise, anoxia, or the injection of adrenalin or serotonin. On the other hand, high levels of lipoproteins tend to

inhibit fibrinolysis (Sherry et al, 1959, MacMillan and Mustard, 1961)

It has been known for decades that normal plasma can neutralize large amounts of thrombin Morawitz (1904) proposed the name *antithrombin* for this activity in plasma The neutralization of thrombin in plasma, however, is a complex action and is not due to a single effect Thrombin can be neutralized without being destroyed Breakdown products of fibrin digestion can do this by blocking the thrombin-fibrinogen reaction Heparin in combination with a plasma factor (heparin-cofactor) acts in a somewhat similar manner There is, however, an enzyme in plasma (true plasma antithrombin) which destroys thrombin, this has been shown to be an α globulin Relatively large quantities of this are present in normal plasma In fact the amount of antithrombin available far exceeds the amount of thrombin formed during coagulation

TESTS EMPLOYED IN THE INVESTIGATION OF DEFECTS IN THE CLOTTING MECHANISM

Coagulation Time. This is the time taken for blood to clot after it is withdrawn from the body Decreased coagulability of the blood is manifested by lengthening of the coagulation time and vice versa A number of methods have been developed to measure the precise moment in which clotting occurs The clotting time of normal blood will depend upon the size of the tube, the number of tubes employed, the temperature at which the blood is kept during the test, and the amount of tilting and agitation of the tube It is important to do the clotting test in exactly the same manner each time If the method is standardized, serial determinations of the clotting time will provide useful information

A modification of the Lee-White method for whole blood clotting time is widely used One-milliliter quantities of blood drawn by clean venipuncture are placed in each of three small test tubes (11-mm internal diameter) placed in a water bath A stop watch is started at the time of venipuncture After 3 min, the first tube is tilted at 15-sec intervals When the blood clots in this tube, the next tube is tilted until clotted and then the third tube is treated similarly. The clotting time is taken as the time of clotting of the blood in the third tube The coagulation time by this method varies between 9 and 11 min for normal individuals By using three tubes, variations in clotting time due to variation in rate and degree of tilting (surface effect) are reduced

In small children venipuncture may be difficult and the clotting time can be measured on capillary blood With a sharp lancet, a drop of blood is obtained from the finger, ear lobe, or heel and allowed to run into a capillary tube The ends of the tube are sealed with plasticine and then the tube is immersed in a water bath at 37° C At 15-sec intervals small fragments of that tube are broken off and the end point is taken when a string of clot is observed The normal coagulation time by this method is between 3 and 5 min, it is shorter than clotting time by the Lee-White method because of the relatively greater surface area of glass in contact with the blood

Prothrombin Time. In the one-stage method developed by Quick, excess of tissue thromboplastin and optimal calcium are added to citrated or oxalated plasma The tubes are placed in a water bath at 37° C and the time for clotting to occur is recorded A standard value for the day is determined using a sample of normal plasma This standard time will vary between 14 and 16 sec A dilution curve may be prepared with normal plasma diluted with adsorbed plasma (prothrombin free) to relate clotting times to prothrombin concentration in per cent of standard This is done in an attempt to reduce the variability caused by the tissue thromboplastin, the activity of which differs from day-to-day It is doubtful, however, whether the dilution curve is worth the effort in the practical management of patients taking oral anticoagulants (Quick, 1945)

Clot Retraction Time. In this test 2 to 3 ml of venous blood are placed in a test tube kept at a 37° C in a water bath After clotting has occurred the clot is freed from the sides of the tube by means of a wooden applicator stick Normally in a couple of hours the clot contracts to a firm mass expressing serum from its interstices Clot retraction is normal in hemophilia but abnormal in thrombocytopenia

Bleeding Time. Bleeding time is determined by nicking the skin with a pointed scalpel blade, the ear lobe being a convenient site. Drops of blood are wiped away every few seconds with filter paper The end point is taken as the moment when bleeding ceases

The normal bleeding time is about 2 to 7 minutes The ability to seal the wound depends on the platelets and on the state of the small blood vessels The bleeding time is normal in hemophilia This at first seems to be a paradox until it is realized that the platelets and blood vessels are normal in this condition In hemophilic patients manipulation of the small wound later causes renewal of bleeding which may be extremely troublesome to control It is believed that this is due to failure of an effective clot to back up the platelet plug (Tocantins, 1936)

Thromboplastin Generation Test. This test resembles the one-stage prothrombin time, but instead of saline extracts from acetone-treated brain as a source of thromboplastin activity, plasma components and platelets are first incubated together to produce plasma thromboplastic activity Normal thromboplastic activity requires factor VIII and factor IX from plasma, cephalin obtained from platelets, and calcium The generation mixture is kept at 37° C and at 1-min intervals thromboplastic activity is tested by adding samples to a substrate of normal platelet-poor citrated plasma and calcium The time of clotting of the substrate reflects the thromboplastic activity of the generation mixture In the test 0 2-ml quantities of normal alumina plasma diluted 1·5 (factor VIII), normal serum diluted 1 10 (factor IX), cephalin or platelets and M/40 calcium chloride solution are placed in a test tube At intervals of 1 min, 0 1 ml of the mixture is withdrawn into a Pasteur pipette and discharged simultaneously with 0 1 ml of M/40 calcium chloride solution into a tube containing 0 1 ml of normal platelet-poor citrated plasma A convenient substitute for platelets is cephalin prepared from brain or soy bean The clotting times of the substrate samples may be plotted as a curve showing normal thromboplastin generation Having prepared a normal thromboplastin generation curve, aluminium plasma or serum from a test subject may be substituted for one of the normal components and the resulting curve compared with the normal If the subject lacks a component necessary for thromboplastin generation, decreased thromboplastic activity will be detected (Biggs and Douglas, 1953)

Partial Thromboplastin Time. This test is widely used and has the advantage of being a one-stage test that is simpler to perform than the thromboplastin generation test It also resembles the Quick one-stage test for prothrombin time Instead of adding acetone-treated brain to plasma cephalin is used in this test If normal amounts of factors VIII and IX are present in the plasma a short clotting time will be observed If one of these factors is lacking, the time will be prolonged Since the early stages of clotting are influenced by surface contact it is customary to add kaolin to the plasma while doing the test to provide maximum activation of contact factors The test may be used for both the diagnosis and control of treatment of hemophilia

Abnormalities of Clotting Process

HEMORRHAGIC DISEASES

Hemophilia. In this condition the coagulation time (but not the bleeding time) is usually prolonged In some cases the blood may not clot for several hours Fatal hemorrhage may follow a minor wound or a simple operation such as a tooth extraction It is important to realize, however, that certain hemophiliacs have a clotting time which is normal or only slightly prolonged In these patients it is necessary to confirm the diagnosis with the thromboplastin generation tests or partial thromboplastin time Hemophilia is transmitted as a sex-linked recessive character On a statistical basis, females carrying the abnormal X chromosome transmit the disease to half their sons while half their daughters will be carriers A hemophilic father will have normal sons but carrier daughters If a hemophilic man marries a hemophilic carrier, it is possible to have a female hemophilic This occurrence is very rare Even hemophilic patients with normal or slightly prolonged clotting times may bleed to death following procedures such as tonsillectomy or tooth extraction. There is, therefore, really no such thing as mild hemophilia Since the disease carries an increased mortality it would eventually disappear were it not for the appearance of fresh mutations Approximately 30% of patients with the disease do not have a previous family history

Two distinct types of hemophilia may be recognized by laboratory testing Certain patients lack factor VIII while others lack factor

IX From the practical standpoint it is important to distinguish between the two types of hemophilia Factor VIII deficient patients require fresh blood or plasma since this activity disappears rapidly on storage Both human and animal factor VIII concentrates are now available Their use has saved many lives and major surgery has been made possible In factor IX deficiency it is not necessary to use fresh plasma since the factor is more stable on storage Satisfactory concentrates of factor IX are not generally available

Afibrinogenemia. The lack of fibrinogen may be either inherited or acquired Because fibrinogen is formed in the liver, in the end stages of serious liver disease the level in the blood may fall It may also temporarily disappear due to destruction by fibrinolysins This may occur following surgery on prostatic cancer, presumably due to release of fibrinolytic enzymes into the blood stream It also may occur in abruptio placentae when sudden and abrupt separation of the placenta may also result in release of fibrinolytic and thromboplastic material into the blood stream and thus, cause a rapid destruction of fibrinogen with bleeding

Purpura. Purpura is hemorrhage into the skin and mucous membranes It may be due to an abnormality of the platelets or of the blood vessels Thrombocytopenic purpura may be either idiopathic or secondary to a toxic agent, the former variety frequently responds very well to splenectomy The mechanism of vascular purpura is not well understood and the diagnosis is made very often by excluding other causes Damage to the blood vessel wall may occur in scurvy, in uremia and in severe infections An important type of vascular purpura is associated with hypersensitivity or allergy

Hypoprothrombinemia. Depression of the prothrombin level in the blood (hypoprothrombinemia) may be found in severe liver damage and in conditions where there is malabsorption of vitamin K, such as obstructive jaundice and steatorrhea Hemorrhagic disease of the newborn results from lack of vitamin K in the mother It may be prevented by giving the mother vitamin K before the baby is born Administration of coumarin drugs results in lowering of the plasma level of prothrombin and clotting factors VII and X

INTRAVASCULAR CLOTTING—THROMBOSIS

Coagulation of blood within vessels may be brought about experimentally by the injection of thrombin or thromboplastin into the blood stream There is an immediate reduction in the number of platelets and the level of fibrinogen After a short time the platelet count returns to normal suggesting that the platelets were not destroyed but were merely sequestered somewhere in the vascular system Return of the fibrinogen level takes longer and is not complete for 24 hr, indicating that fibrinogen was destroyed and was replaced by newly formed fibrinogen

Injury to the vessel wall results in a deposition of platelets from the flowing blood at the site of injury The platelets stick to the injured area and also to each other forming a platelet plug On top of the platelet, nidus fibrin threads are formed which enmesh red and white blood cells This is known as a *thrombus*, and it is important to distinguish this from a simple blood clot formed in a stagnant pool of blood A thrombus may either partially or completely plug the lumen of a vessel The "head" of a thrombus at its point of attachment consists mainly of platelets and fibrin The "tail" of the thrombus frequently lies within the lumen of a vessel and consists of fibrin and red cells

A thrombus or a portion of it may break free and being carried away in the blood stream may lodge in some remote area The portion of the thrombus that travels is known as an *embolus* Emboli from the leg veins lodge in the lung, producing an area of consolidation in the lung known as a pulmonary *infarct* If an embolus originates in the left side of the heart it will be carried to a peripheral artery, for example, a cerebral artery giving rise to a cerebral infarct

Causes of Thrombosis in Man. The causes of thrombosis in man may be considered under the following headings

Injury to a Blood Vessel Injury to a blood vessel by trauma, or the application of an irritating substance leads to thrombosis, but the basic mechanism underlying its development remains obscure It is possible that a damaged endothelial lining causes the release or activation of substances that initiate the clotting process

Deposition of Platelets Dietrich (1919) some years ago made an experimental study of postoperative thrombosis He described the deposition of the platelets and compared them with sand which is kept in suspension in a rapid stream Just as sand is deposited in a ribbed pattern upon the seashore or at a river's mouth where the current is slowed, so platelet deposited in a thrombus form characteristic ridges The identification of platelet strands or ridges enables the pathologist to distinguish a thrombus from a postmortem clot In vitro the addition of adenosine diphosphate (ADP) to a platelet suspension causes rapid clumping of the platelets It is possible that damaged endothelium releases substances similar to ADP which alter the platelet membrane causing them to adhere to each other and to the endothelium The platelet mass sometimes completely obstructs the vessel Occasionally it breaks away and forms a small embolus When this happens, another plug of platelets forms at the site of injury (Wright, 1942, Gaarder et al, 1961, Moolten et al, 1949)

Slowing of the Blood Stream After major surgery or childbirth, there is an increased risk of developing thrombosis and embolism This may be due to the fact that the flow of blood in the veins becomes sluggish, with a resulting platelet deposition and clotting Early ambulation after surgery and childbirth has been accompanied by a lowered incidence of thrombosis

Changes in the Blood After operation and childbirth, both the number of platelets and the level of fibrinogen are increased Many investigators have carried out tests of clotting factors in an attempt to demonstrate a prethrombotic state All such studies have been inconclusive, while some workers have reported changes, these have not been confirmed by other investigators An important factor leading to thrombosis is probably an alteration in platelet stickiness, associated with alterations in the endothelium and slowing of blood flow

Summary. The first event in the development of a thrombus is probably an alteration in the endothelial lining Platelets adhere to this area and to each other forming a white thrombus The growth of the thrombus occurs by the deposition of further platelets and the formation of fibrin, which enmeshes red and white cells The accumulation is favored by a sluggish blood flow The clot that has been formed may remain fixed and either completely or partially occlude the blood vessel The tail of the thrombus may be free in the flowing stream of blood The thrombus may grow by the fresh deposition of platelets and fibrin The thrombus may have a precarious hold on the lining of the vessel and break loose from its point of attachment, or a piece of the tail may become detached When this occurs an embolus is carried in the blood stream to lodge in a distant part Once the thrombotic process is recognized in leg veins, treatment consists of moderately active exercise, e g, walking to increase venous return from the part and to lessen the change of a clot's forming in stagnant blood Secondly, anticoagulant therapy with heparin or dicumarol may be given with the hope of lessening the buildup of a thrombus by fresh clot Anticoagulant therapy might be expected as well to allow the natural fibrinolytic process to proceed without hindrance

PREVENTION OF CLOTTING

The Means Used to Prevent or Retard Coagulation—Anticoagulants. *Cold* Since the clotting mechanism consists of a series of chemical and enzymatic reactions, it is to be expected that coagulation will be retarded by lowering the temperature Keeping blood at a temperature of from 5 to 10° C postpones coagulation but does not absolutely prevent its occurrence Cold, e g, ice, etc, applied to the surface of the body as a means of arresting hemorrhage has, however, no effect in retarding the coagulation process Actually a hemostatic action is exerted in such instances as a result of the vasoconstriction reflexly induced

Avoidance of Contact of the Blood with Foreign Materials or Injured Tissues Drawing the blood directly from an artery or vein (in order to prevent contact with the tissues—a source of thromboplastin), and collecting it into a receptical whose walls have been made smooth and *unwettable* by a coating of paraffin, or preferably of *silicone*, will slow greatly the coagulation process Silicone is the most effective substance known for this purpose, it acts like the vascular endothelium in inhibiting the breakdown of platelets (Jaques and associates,

1946) and the formation of thromboplastin activity Plasma, after it is collected in siliconelined apparatus and centrifuged at high speed in order to remove the platelets as completely as possible, will remain fluid for 3 or 4 days even at room temperature The importance of a smooth surface in maintaining the fluidity of blood is also shown clearly in the classical experiment of the "living test tube " If a section of the jugular vein of an animal, e g , the horse, be isolated between ligatures and carefully removed with its contained blood, this will not clot for a long period, not until changes occur in the lining of the vessel wall and the necessary thromboplastin is thereby provided

Decalcification The addition of oxalate (sodium, potassium, or ammonium) or a fluoride to blood to the extent of 0 1% or more, completely destroys its power to clot spontaneously In the former instance the calcium is precipitated as calcium oxalate Oxalated blood recovers its ability to clot if shaken with chloroform, but the fibrin which forms redissolves after the blood has stood for a time Fluoride does not precipitate the calcium but forms a weakly dissociated calcium compound Sodium citrate also prevents coagulation, in this instance a double salt—calcium sodium citrate—is formed, which again is only slightly dissociated

Neutral Salts Magnesium sulfate solution in a strength of 27% mixed in the proportion of 1 4 of blood postpones coagulation for some time, but does not prevent it indefinitely Sodium sulfate in half-saturated solution added to an equal quantity of blood has a similar effect, as has also a 10% solution of sodium chloride in the same proportion The manner in which these substances act is not clear They do not decalcify the blood

Anticoagulants Used Therapeutically. At the present time there are two types of anticoagulants used for therapeutic purposes The anticoagulant heparin has an immediate direct action on the coagulation system, whereas, a group of drugs represented by dicumarol has an indirect effect on the coagulation system through their action on the liver

Heparin is a naturally occurring sulfonated polysaccharide with strong anticoagulant action It can be prepared from bovine lung on porcine or bovine intestinal mucosa It requires a plasma fraction for its antithrombic activity but can inhibit at a number of points during the coagulation process (Fig 4 30) Heparin must be given parenterally since it is not absorbed from the alimentary tract Because it is rapidly inactivated and removed from the circulation, it is difficult to maintain a therapeutic level in the blood while avoiding the danger from bleeding Prolonged action can be obtained by intramuscular injections of concentrated solutions of heparin Heparin readily combines with basic peptides and proteins and can be inactivated by protamine

Dicumarol (4-hydroxycoumarin), which was originally extracted from spoiled sweet clover, lowers the level of prothrombin and associated coagulant factors by acting as an antagonist to vitamin K (Link, 1944) There is now a large number of analogues of dicumarol and phenolindane-dione compounds which are more potent than dicumarol but act similarly The effect of these drugs can be prevented or overcome by the administration of vitamin K These drugs have no effect on coagulation when added directly to blood

The whole problem of anticoagulant therapy is undergoing a great deal of study and there are many physicians who question the therapeutic value of anticoagulants This problem, however, is beyond the scope of this chapter

SUBSTANCES WHICH HASTEN THE CLOTTING PROCESS (HEMOSTATICS)

It has been widely accepted that epinephrine hastens coagulation, but such an effect has not been proved The effects of thrombin and thromboplastin have been discussed Tissue extracts, especially those of lung and thymus which are rich in thromboplastin, are powerfully coagulant, as are also the venoms of some species of snakes The coagulant property of such venoms depends upon their containing a proteolytic enzyme that converts prothrombin to thrombin, the conversion occurs in the absence of ionized calcium

Thrombin sprayed upon the bleeding surface in conjunction with fibrinogen forms a covering film or foam of fibrin, and is an effective means of arresting bleeding, especially from numerous small vessels. Among other hemostatic agents used to hasten the clotting process are sodium alginate, derived from seaweed, and a gauze made of oxidized cellulose, which swells when soaked with blood When sodium alginate comes into contact with the blood and serum of the wound, it is converted to calcium alginate, which "clots" to form a tenacious layer

4. THE LYMPHATICS, LYMPH, AND TISSUE FLUIDS

General Considerations

STRUCTURE OF THE LYMPHATIC SYSTEM

The lymphatics can not be discussed without reference to blood capillaries (sect. 3, chap. 1). Their very close interrelation is shown in Figure 4-31. The part of a capillary nearest to the arteriole is called an arteriole capillary and that part nearest to the venule is designated as venous capillary. There are "through" capillaries in which blood flows continually, and there are branched capillaries in which the blood flow is regulated by vasoconstriction and vasodilatation mediated through the precapillary sphincters. The wall of the capillary consists of a single layer of flat endothelial cells resting on a basement membrane and surrounded by a thin adventitia of fibers and cells. The capillary endothelium, when appropriately stained, looks like pavement blocks.

The endothelial cells of blood capillaries have two distinct ultrastructural characteristics. *a*) The process of pinocytosis[1] takes place. This can be considered to be microphagocytosis. One can observe tiny pinocytic vesicules that appear by invagination of cell membranes and enclose fluid and sometimes particulate material. The direction of transport across capillaries is primarily into tissue at the arterial capillaries and into the capillaries at the venous end. *b*) In addition the capillaries in some anatomic areas appear to have pores varying in diameter from 300 to 500 Å, usually with a diaphragm overlying them. The pores play a role in the movement of fluid and solutes through the capillaries, and are particularly prominent in the glomerular capillaries. The role of osmotic pressure and hydrostatic pressure in controlling the flow of fluids out of the vascular system is discussed elsewhere (sect. 3, chap. 1). In Figure 4-31 the knoblike ends of lymphatic capillaries are shown almost in contact with blood capillaries. The direction of flow of fluids and solutes out of the arterial capillary and into tissue spaces and back into the venous capillaries and into lymphatic capillaries is shown by the arrows.

By combining the injection of colloidal materials and electron microscopy fundamental observations have been made upon the structure and function of the lymphatic capillary (Leak and Burke, 1966, 1968; and Cliff and Nicoll, 1970). The endothelium of the lymphatic capillaries is similar to that of the blood capillaries. There is no basement membrane. The diameters of the lymphatic capillaries vary widely from dilated terminal capillary bulbs to structures that are either very narrow and slitlike, or dilated capillaries, depending upon their state of function. Pinocytosis has been clearly demonstrated, it is one method of transportation of material from the tissue spaces into the lymph capillaries. The electron micrographic studies of the lymphatic endothelium show that in some areas away from the nucleus the endothelium is very attenuated, being in some places little more than the two membranes in width. The cytoplasm of the endothelial cells is either interdigitated or overlapping, although in some instances there appear to be wide gaps between the cells. In addition, there are collagenous anchoring filaments attached to lymphatic endothelium and then extending out to unite with collagenous bundles in the adjoining tissue. The detail of this structure is shown in Figure 4-32. When fluid accumulates in the interstitial spaces by transudation (or is injected, one can visualize that this fluid will occupy space between the lymphatic capillary and tissue collagenous fibers. Since the fibers attached to the lymphatic endothelium and the tissue collagen are inelastic one would expect the over-

[1] The absorption of liquids and tiny particulates by cells—"to drink."

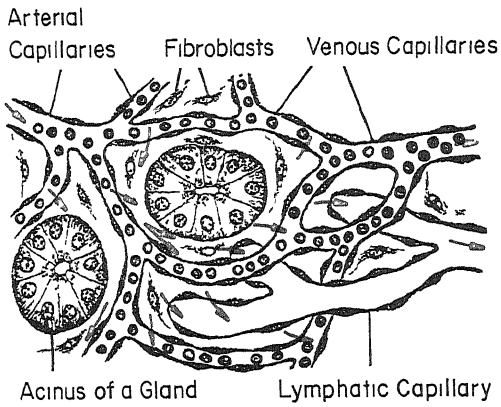


Fig 4 31 Diagram showing the relationship of lymphatic capillaries to blood capillaries and to the tissue fluids around the acini of a gland Similar relationships exist in most of the organs of the body Arrows indicate the direction of flow of fluid leaving the arterial capillaries permeating the connective tissue spaces as tissue fluid and re-entering the blood capillaries on the venous side The lymphatic capillaries supplement the venous capillaries in the drainage of fluid from the tissues to the circulatory system (From Copenhaver 1967)

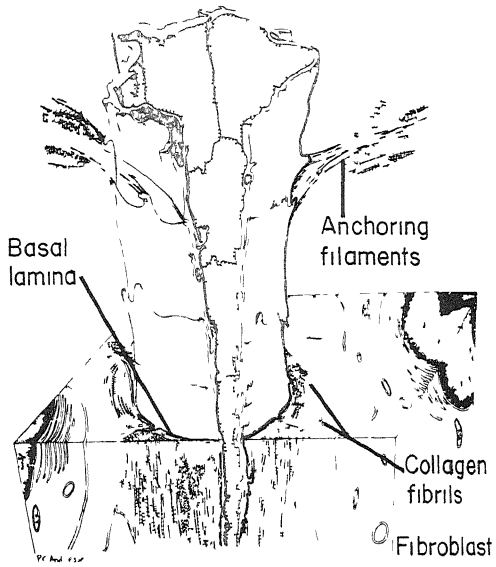


Fig 4 32 A three-dimensional interpretative diagram of lymphatic capillary that was reconstructed from collated electron micrographs The three-dimensional relation of the lymphatic capillary to the surrounding connective tissue area is illustrated The lymphatic anchoring filaments appear to originate from the endothelial cells, and extend among collagen bundles elastic fibers and cells of the adjoining tissue area thus providing a firm connection between the lymphatic capillary wall and the surrounding connective tissue Irregular basal lamina and collagen fibers are as marked (From Leak and Burke 1968)

lapping endothelium to be pulled apart so as to provide large pores through which the edema fluid under greater pressure can flow into the lumen of the lymphatic capillary The combination of a mechanism to open up the spaces between endothelial cells and an active process of pinocytosis gives two mechanisms by which fluid and its solutes can be rapidly transported from blood capillaries through the interstitial spaces and then into the lymph capillaries In the bat wing Cliff and Nicoll (1970) have shown that the terminal lymphatic capillary bulbs enter into collecting ducts that possess a contractile wall and valves These emerge into transport lymphatic channels with very active contractility and valves The transport lymphatics combine into larger and larger lymphatic vessels propelling the lymph to and through lymph nodes and toward the great veins Ultimately major lymphatic vessels are formed in the extremities and within the viscera, and then merge into the large right lymphatic and thoracic ducts From these lymph pours into the blood stream by way of the right and left subclavian veins, respectively

In the course of flowing from the periphery to the entrance in the blood vessels, the lymph flows through one or more lymph nodes As the lymphatics approach the node, they break up into finer channels called the afferent lymphatics, These penetrate into the sinuses of the cortex of the lymph node, then through the cortex and the medullary sinuses which are lined by phagocytic cells One or more large efferent lymphatics formed by confluence of the numerous medullary lymphatics emerge from the hilum of the lymph node, and progress onwards to join the major lymphatic vessels

In the walls of the abdominal cavity the lymphatics are most abundant on the under surface of the diaphragm, where the greatest lymphatic absorption takes place from the peritoneal cavity There is also substantial absorption into the lymphatics of the omentum Lymphatic systems exist in all organs with the exception of the central nervous system, cornea and the alveoli of the lungs Protein is not readily absorbed from the interior of the alveoli into the lymph capillaries because of the poor lymphatic supply, or possibly because it cannot readily penetrate the alveolar wall Water passes readily into the lung capillaries, however, because of the low hydrostatic pressure within them

THE FLOW OF WATER ELECTROLYTES SMALL MOLECULES AND PROTEINS BETWEEN THE BLOOD AND LYMPH

The rapid flow of oxygen and nutrients to cells and the return of carbon dioxide and metabolic waste products is accomplished because the blood capillary membrane is freely permeable to water and crystalloids There is a rapid and continual change between the intravascular and extravascular compartments by diffusion The transcapillary exchange takes place on an exceptionally large scale without any perturbation to the circulating blood volume or the electrolyte balance The magnitude of this flow is astonishing Once every minute a quantity equivalent to the total plasma sodium chloride and water are exchanged across the blood capillary membrane (Flexner et al , 1942, Merrell et al 1944, Chinard et al , 1955) This bulk movement of fluid into and out of the blood capillaries depends on the balance of hydrostatic and osmotic pressures acting across the capillary endothelium The essential characteristic of a capillary wall that regulates the net movement of fluid is its relative impermeability to plasma protein molecules The osmotic force exerted by the circulating plasma protein molecules within the blood capillaries counterbalances the hydrostatic pressure in the capillaries The capillary wall, however, is not completely impermeable to plasma proteins, and all protein molecules leak through to a certain extent Nevertheless, they are found in the interstitial fluid and lymph in significantly lower concentration than in plasma (Wasserman and Mayerson, 1952, Yoffey and Courtice, 1956) The accumulation of proteins in the interstitial fluid, with an increasing osmotic pressure that would disrupt the balance of forces controlling the exchange of fluid across the capillary membrane, is precluded by the rapid flow of protein molecules directly into the lymphatics, presumably by a combination of pinocytosis and flow through the gaps in the lymphatic capillary endothelium

The escape of plasma proteins into interstitial tissues varies with the tissue, it is greatest in the liver Factors that control the escape of protein into the interstitial space are not easily described by any simple concept It has been shown that a quantity roughly 50 to 100% of the circulating plasma proteins escapes across the blood capillary membrane and re-enters the blood through the lymphatics each day (Yoffey and Courtice 1956)

Composition and Flow of Lymph

The *protein concentration* of lymph is always lower than that of plasma, even though there is considerable variation between lymphatics depending upon the organ As oral or intravenous fluid intake is increased the protein concentration of lymph decreases

Lymph contains all of the *coagulation factors*, and clots although less readily than blood plasma There is a much higher concentration of some coagulation factors in the hepatic lymph in contrast to peripheral lymph because these factors are made within the liver Antibodies are found in the lymph and will be discussed in more detail in Chapter 10 Almost all of the enzymes that are found in plasma are also found in lymph to a lesser extent

The *electrolyte* concentration in lymph is not essentially different from that of plasma The total cation is slightly lower in the lymph than in the plasma and chloride and bicarbonate levels tend to be higher The direction of these differences in concentration is consistent with the Gibbs-Donnan equilibrium operating on two phases whose concentrations of nondiffusible ions (proteins) differ

Of the plasma *lipids*, cholesterol and phospholipid, being mainly associated with protein as lipoprotein, are present in the lymph in concentrations which vary with the level of protein in the lymph (Yoffey and Courtice, 1956) Neutral fat in the form of chylomicrons depends on the degree of fat absorption from the alimentary tract In the postprandial state one finds large quantities of lipoproteins and fats in the lymph coming from the gastrointestinal tract Between meals, the fat in the thoracic duct drops to low levels

The *other constituents* of the plasma that are readily diffusible, nonprotein and nonelectrolytes are present in the lymph in concentrations approximately the same as in the plasma After a carbohydrate meal when the blood sugar rises rapidly or after intravenous injection of glucose, the lymph sugar lags behind but soon an equilibrium is established between the plasma and lymph The nonprotein nitrogen, urea, amino acid, and creatinine levels are nearly the same in plasma and

lymph Small differences observed probably arise from errors of the techniques

The lymph contains *cells* Lymphocytes of all sizes and degree of maturity are the most numerous A very rare monocyte and macrophage is found Platelets are not observed Red cells, when present, indicate the degree of bleeding into the tissues as a result of injury, or when found in the thoracic duct suggest the presence of intestinal parasites Granulocytes are found in the lymph draining areas of infection Plasma cells are also found in very small numbers The concentration of lymphocytes in efferent lymph leaving a lymph node is always much greater than that in afferent lymph The circulation of the lymphocyte, its proliferation and factors that influence this will be considered in Chapter 10, along with the structure of lymph nodes and their role in immunity and defense against infection

THE FORMATION PRESSURE, AND FLOW OF LYMPH

Any condition that increases the outpouring of fluid from the capillaries into the tissues will increase the flow of lymph if there is no obstruction The factors influencing the flow are the contractility of the lymph vessels, activity of the skeletal muscles, peristalsis, and a differential pressure between interstitial spaces and lymphatics The average tissue pressure is about 1 9 cm of water, whereas, the pressure in the lumen of the lymphatic capillary is 1 2 cm of water (McMaster, 1947) Tissue fluid, therefore, enters lymphatic capillaries because of the gradient of pressure and through the gaps in the lymphatic capillary endothelium described earlier Presumably, the retention of fluid inside of the lymph capillary is in part due to the higher pressure of the interstitial fluid outside Propulsion of lymph from one region to another is brought about by the extrinsic and intrinsic forces described earlier The lymph pressure in the vessels will rise and the lymph will move centripetally to a region where the pressure is lower The numerous valves within the lymphatics prevent a retrograde flow of the lymph Since lymph flows centripetally along a duct there must be a gradient in pressure. During skeletal muscular contraction the lymphatics empty as tissue pressure increases, and during relaxation the vessels fill again.

FUNCTIONS OF THE LYMPHATIC SYSTEM

A most important function is obviously the return of protein, water and electrolytes from tissue spaces to the blood The lymphatics are exceptionally important in the absorption of nutrients, particularly fats from the gastrointestinal tract

The lymphatics serve also as a transport mechanism to remove red blood cells that have been lost into the tissues as the result of hemorrhage, or bacteria that may have invaded the tissues When there is an infection in a distant part of the body, the regional lymph node becomes inflamed as a result of the localization of bacteria or toxins carried in the lymph to the gland The lymph nodes contain a very efficient filtration system in the cortical and medullary sinuses, these are lined with phago cytic cells that engulf bacteria and red cells or other particulate material The efficiency of the filtration system can be demonstrated by the direct perfusion of pathogenic bacteria into lymphatics that are afferent to a lymph node Simultaneous culturing of the efferent lymphatic will show it to be sterile The capability of the lymph node to filter out pathogenic bacteria can be overcome, however, and bacteria escaping through the efferent lymphatics into the blood may produce a septicemia

CONDITIONS THAT INCREASE THE LYMPH FLOW

1) Increase in Capillary Pressure Landis and Gibbon (1935) found that filtration from the capillaries shows a definite increase when the venous pressure rises above 12 to 15 cm of water The rate of filtration from the capillaries is directly proportional to the increase in venous pressure (Fig 4 33) In their experiments, when the venous pressure was increased to any given level the filtration rate increased rapidly at first, but gradually slowed and finally ceased This falling off in the filtration rate they ascribed to a rise in extracapillary pressure due to the fluid accumulation, which opposed the hydrostatic pressure within the blood capillary The accumulation of extracellular fluid may be expected therefore to be greater in regions that are loose in texture and where the skin is readily stretched In persons with firm, resistant skin, edema, for the same reasons, is later in appearing and less pronounced

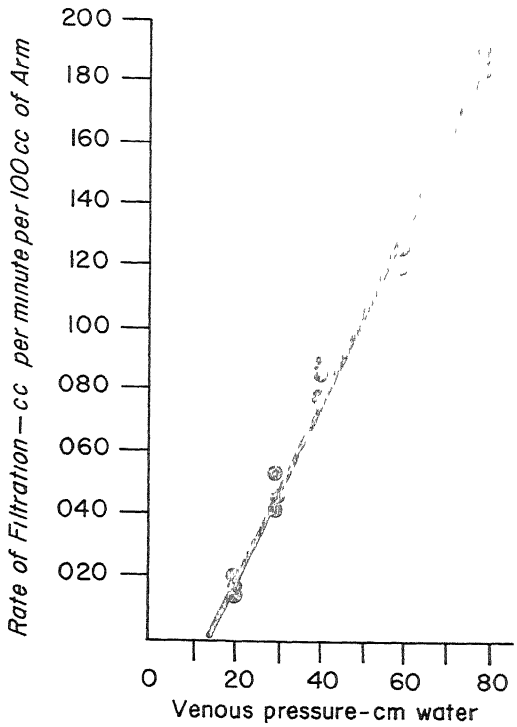


Fig 4 33 Rates of filtration produced during 30 min by venous pressures between 20 and 80 cm of water (After Landis and Gibbon 1933)

Increased pressure in the portal vein or in the hepatic veins, produced for example by obstruction, causes increased filtration into the tissues of the abdominal viscera and a great increase in the volume of lymph flowing through the thoracic duct Increase in arterial pressure, on the other hand, does not increase filtration in animals until the pressure reaches around 300 mm Hg

2) Increase in Capillary Surface area. This increases the leakage of protein and fluid It may follow any change that causes distension of the capillary vessels, e g., *a)* increase in capillary pressure, b) increase in the local temperature, or c) infusion of fluid

3) Increase in Capillary Permeability.

a) Rise in temperature locally may increase capillary permeability

b) Capillary poisons such as peptone increase the flow of lymph from the thoracic duct, probably as a result of its injurious effect upon the abdominal capillaries The increased flow occurs after removal of the liver, so injury to the hepatic vessels is not essential Other substances that increase lymph flow in this way are extracts to strawberries crayfish mussels leeches, histamine, and foreign proteins To what extent this is due to a change in permeability and to what extent it is a consequence of capillary dilation is not known

c) Reduced oxygen supply to the tissues increases lymph flow, probably because of dilation of blood vessels but possibly also through damage to the capillary endothelium

4) Hypertonic Solutions The intravenous injection of a concentrated solution of glucose sodium sulfate, or sodium chloride causes an increased flow of lymph from the thoracic duct These substances in concentrated solution may increase permeability of the capillary wall but also exert an osmotic effect that alters the normal equilibrium between the extravascular and intravascular fluids Water at first enters the plasma from the tissue spaces, particularly of the muscles and subcutaneous tissues of the limbs, which in consequence show a fall in volume, the brain shrinks Thus removal of fluid may actually extend to the fluids within the cells which undergo shrinkage, general desiccation of the tissues may result The blood volume for the time is greatly augmented, the excess fluid for the most part being accommodated in the capacious capillary and venous areas of the abdomen As the electrolyte injected moves out of the plasma and into the interstitial fluid it carries water with it The viscera—liver, kidneys, spleen, and intestines—then increase in volume, due not only to the distention of their vascular beds, but more importantly to the great outpouring of fluid as well as of protein that occurs through the capillary walls This fluid then swells the volume of lymph in the thoracic duct In this way these substances produce a redistribution of fluid

The injection of isotonic saline will also increase the passage of protein and fluid from the capillaries and increase lymph flow This too is probably a consequence of the changes in osmotic and hydrostatic pressures in the capillaries

5) Increased Function Activity. When a gland or muscle enters into activity an increase in lymph flow occurs which starts a little after the commencement of the secretory or contractile response, but is nearly synchronous with the increased metabolism resulting from the activity The increased flow is as-

cribed to *a*) formation of metabolites that increase the osmotic effect of the tissue fluids and cause more fluid to leave the vessels, *b*) vasodilation, with increased capillary pressure and increased fluid and protein leakage

During rest the flow along the lymph vessels of the muscles and subcutaneous tissues is slight, and the protein content of the lymph is high During activity the protein concentrations fall since less transuded water undergoes reabsorption into the blood and more is carried away by the lymph channels This may occur even though the leakage of protein from the vessels is greater The contracting muscles exert a pumping effect upon the lymph driving it along the vessels

6) Massage and Passive Movement. These act to a certain extent like muscular activity They augment the blood flow, capillary pressure, and capillary surface and so increase lymph formation The manipulations and movements of the muscles serve to propel the lymph along the lymphatic channels

Edema

Edema is a term applied to an excessive accumulation of fluids in the tissue spaces, and is due to a disturbance in the mechanisms of fluid interchange, which have been considered in the preceding pages Instead of there being a perfect balance between the inward and outward flow of fluid through the capillary membrane, absorption is exceeded by transudation The particular factor or factors of the mechanism that are disordered are not always clear, and a satisfactory explanation of all forms of edema cannot be given But from previous discussions it is evident that the following factors will tend to increase the volume of interstitial fluid *a*) reduction in the protein concentration of plasma, (edema commences when the albumin fraction has fallen to between 2 5 and 3%), *b*) a general or a local rise in capillary blood pressure, *c*) increased permeability of the capillary membrane, *d*) increase in the filtering surface as when the capillaries dilate, and *e*) obstruction of the lymph channels

There is a tendency for the accumulation of edema fluid to progress so far and then become stationary, provided that the conditions producing it remain constant, for when the tissue fluid pressure reaches a critical level its opposition to the force driving fluid from the vessels prevents further transudation It should be appreciated too that as a source of edema fluid the plasma has definite limitations in a quantitative sense, and unless the total extra cellular fluid is increased a generalized edema cannot be very severe

Since edema is only a symptom of some primary condition it may have a variety of causes, according to the particular disease with which it is associated

1) Cardiac Edema. In congestive heart failure there is both an increase in extracel lular fluid and a salt retention The volume of this extracellular fluid can be reduced by re striction of the salt intake or by the administration of diuretics that remove both water and salt Or it can be discharged by improving cardiac action by digitalis

2) Mechanical Obstruction of Veins When the main veins leading from a part are obstructed by any mechanism, as in cirrhosis of the liver, thrombosis, etc , an increased transudation of fluid occurs This is due in part to the rise in intracapillary pressure and an increase in the filtering surface, but the permeability of the capillary wall may also be increased as a result of the impaired blood supply or, as in the case of tumors, probably by the production of toxic substances as well

3) Edema Due to Renal Disease In chronic nephritis edema is not usually pronounced unless the heart is failing, however, in the nephrotic syndrome it is an outstanding feature In this condition a reduction in plasma protein as a consequence of the loss of protein in the urine leads to the passage of an abnormally large volume of fluid from the capillaries throughout the body This in turn is probably responsible in some way for renal retention of salt and as a consequence of this, retention of water

4) Inflammatory Edema. In this type sev eral factors combine to produce the fluid infiltration of the tissues Increased capillary pressure occurs, due to dilation of the vessels and local slowing of the blood stream as well as to thrombosis and obstruction of the returning veins There is an increase in the filtering surface too, and the lymphatics for a variable distance from the inflammatory area are obstructed as well The capillary walls are also seriously damaged by the bacterial toxin or other injurious agent, so that a fluid with a high protein content escapes from the vessels The edema is localized to an area of varying

extent surrounding the injured site Certain analgesics, e g , aspirin, morphine and Demerol, tend to reduce the edema of simple inflammation

5) **Giant Edema** This is a localized noninflammatory edema that comes on with great rapidity and involves the loose areolar tissue in regions such as the hands face, external genitalia or larynx It has a familial trend Little is known definitely regarding the mode of its production A histamine-like substance liberated at the site of the edema is apparently the immediate cause The remote exciting cause is frequently a foreign protein consumed in the diet, which apparently gains access to the blood stream in a more or less unchanged state, for the attack often follows a particular food to which the subject is susceptible, and is sometimes accompanied by gastrointestinal disturbances The effects are therefore of an anaphylactoid nature and constitute one type of allergy

6) **Edema Caused by Malnutrition or Toxic Substances** Edema may occur in the anemias or in conditions in which the general nutrition of the body suffers When the diet is deficient in vitamins, or there is too little fat or protein in the diet edema may occur, as in beriberi, scurvy, "war edema," or in the faulty nutrition of infants In animals edematous conditions have actually been induced by general underfeeding, or by a diet deficient in fat and soluble vitamins, or by one deficient in protein alone The factors responsible for the increased transudation in these cases are not always clear, but in others there is a marked lowering of plasma protein (more particularly the albumin), which alone is sufficient to account for the edema

Certain chemical substances such as arsenic, salts of heavy metals, and the toxins of certain infectious diseases, such as diphtheria, acute nephritis, etc , are known to act as capillary poisons and apparently cause edema in this way An interesting type of toxic edema is that which may be produced in animals by the injection of hematoporphyrin This substance sensitizes the tissues to ultraviolet rays, and the edema occurs only after the exposure Histamine causes local edema at the point of injection by inducing capillary dilation and increased permeability of the membrane.

7) **Chronic Lymphatic Obstruction.** Widespread obstruction of the lymph vessels may result from congenital or familial disorders of the lymph vessels Lymphedema of the familial type is called Milroy's disease Acquired lymphedema results from obstruction of the lymph channels by neoplasm, scars operative removal of lymph nodes and fibrosis caused by x ray therapy It may also follow a low grade lymphangitis from filariasis from lymphogranuloma venereum or from repeated streptococcal infections The filarial parasite finds its way into the lymph vessels of the limbs and blocks their lumens with the production of the condition known as elephantiasis

Severe edema may occur in the foot and ankle after fracture dislocations, and extensive hemorrhage due to injury to lymphatic capillaries It may be months particularly in older persons, before lymphatic drainage is reestablished Compression bandages enhance lymphatic drainage

For the absorption of fluid the pleural cavities depend upon the lymph channels, the accumulations of fluid may occur here as a result of lymphatic obstruction The edema associated with carcinoma is due chiefly to the filling of the lymphatic channels with cords of cancer cells, as well as to venous obstruction caused by the pressure of the growth The "milk leg" of the puerperium is in part due to lymphatic obstruction The tissue fluid in these types of edema has a relatively high concentration of protein

In the early stages, lymphedema cannot be distinguished from any other form of soft pitting edema On examination of the fluid, the high protein concentration separates it from cardiac and nephritic edema but not from the fluid of myxedema Ultimately, lymphedema causes fibrosis in the tissues, and in time the tissues become hard and brawny The skin may be thick and folded with indulent ulcers

8) **Heat Edema.** The effect of heat upon capillary permeability has been mentioned Excessive heat may actually lead to edema in man It occurs in the tropics and occasionally in so-called temperate zones during intense heat wave It is usually greater in dependent portions of the body Increase in blood volume, enlargement of the filtering surface as a result of the opening up of fresh capillaries, and the rise in capillary pressure incident to the dilation of capillaries previously patent are also factors in the production of this type of edema

10 / LYMPHORETICULAR SYSTEM AND THE LYMPHOCYTE

The lymphoreticular system includes the thymus, lymph nodes, spleen, lymphoid tissue of the alimentary tract, blood and tissue lymphocytes, the bone marrow lymphocytes, and the ceaseless traffic proceeding among these tissues through the blood and lymphatics Rich (1936) stated that ignorance about function of the lymphocyte was disgraceful because pathologists had known for over a century about the presence of lymphocytes in the sites of chronic infections, the large size of lymph nodes during immunization and in the presence of local infections The vast number of lymphocytes in the body indicated a vital albeit unknown function Yet precise knowledge was lacking, and speculation was based on concepts such as possible association of lymphocytes with immunological processes In the interval since Rich's comment, a plethora of techniques has developed for the study of lymphocytes and their relationship to immunity, with the result that knowledge is increasing exponentially Before proceeding to the cellular events connected with immunological processes a discussion of immunological responses and theories of immunity is required

Immunological Responses

TYPES OF RESPONSE

Two general types of immunological responses are recognized

1) The first produces humoral antibodies (immunoglobulins) Classically this type of response has been divided into the *primary response* in which a low but detectible titer of antibody is produced, and the *secondary response* which comes from introduction of the same antigen 2 to 3 weeks later At this time the antigen produces an exponential rise in the serum antibody concentration, and at the cellular level is accompanied by rapid development of germinal centers in lymph nodes and spleen, with intense cellular proliferation (Cottier et al , 1964a) The notion that a primary response represents a first contact with an antigen must be questioned In fact, the idea of a "virgin" immune response is questionable and difficult to rationalize with current thinking Nevertheless, the existence of phenomena called primary and secondary responses is not questioned These phenomena are probably best explained as a result of antigen dosage and timing that determine the number of cells recruited into antibody synthesis

Antibody can be found free in the serum (as well as within cells of all stages of maturation from small lymphocyte to the fully developed plasma cell) Although the appearance of antibodies in high concentration in plasma during the course of an infection had been considered beneficial, the use of antibody-containing antisera in patients with overwhelming infection was disappointing For example, in preantibiotic days it was quite evident that blood cultures became gratifyingly negative after large quantities of antipneumoccocal sera were given intravenously to patients with lobar pneumonia Alas, the patients frequently died This lack of therapeutic benefit has been attributed to the fact that in pneumonia large areas of the lung are "consolidated," with relatively little circulation through the affected region In addition, the immunoglobulins are large molecules and will not diffuse through the capillary endothelium unless there is a marked inflammatory response When they are released directly into the tissues, by contrast, the diffusion distances will be short It has been proposed that a major function of those cells that contain antibodies is the transport of specific antibody by migration of cells to anatomic sites where antibody is needed and to which it can not diffuse readily as a large molecule (Hall, 1969)

2) The second type of immunological response is mediated by cells, serum antibody is apparently incidental In these reactions, cells

of a lymphocytic family develop as part of an immunological reaction, later make direct contact with the antigenic material and then in various ways lead to its destruction or rejection. Examples are delayed hypersensitivity, allograft rejection and immunological surveillance for mutant cells (cancer, autoimmune clones, etc.) harmful to the host.

Humoral and all types of cellular immunity involve recognition of antigen by cells, followed by cellular differentiation and proliferation to amplify the reaction. The lymphocytes are unquestionally involved and in the center of immune responses. An understanding of the migration pathways, lifespan, proliferation rates, pool sizes and fluxes of lymphocytes between blood and the diverse tissues and the factors regulating the population size of these pools is crucial to an understanding of humoral and cellular immunity.

THEORIES OF THE CELLULAR BASIS OF IMMUNE RESPONSE

The instructive theory of antibody formation assumes that there are cells capable of responding to a huge variety of antigenic substances by differentiation and proliferation that produce a family of cells that synthesize the antibody complementary to the antigen. This view is called "instructive" because it postulates that the antigen is providing the cells with information and instructs them to produce the particular antibody. The concept can be presented, not in terms of a "template" idea (Pauling, 1940), but rather in molecular biological parlance. Thus, the antigen may be derepressing the genomes carrying the code for the immunoglobulin complementary to the antigen. The code is transcribed. The messenger-ribonucleic acid (m-RNA) passes into the cytoplasm and its message is translated into the antibody by the protein synthesizing machinery. This concept requires that all lymphocytes be able to respond to any antigen, this, however, does not appear to be compatible with the present information.

A second theory of antibody production is the *clonal selection theory* (Burnet, 1959). This theory envisions that during cell replication there develops a population of progenitor cells heterogeneous in respect to their capacity to respond to various antigens. This heterogeneity is believed to arise from somatic mutations, which affect the cellular genomes so that each member of the clone can respond to a specific antigen. This heterogeneity is assumed to develop by mutation without the intervention of a specific antigen.

A conservative but not established system to explain immunological behavior might embrace the following features. The antigen must be recognized. The simplest method for cellular recognition of antigen seems to be the possession of a coating of complementary antibody to trap the antigen as it passes by the cell, or to trap the cell as it passes by the antigen deposit. Immunoglobulin coating of some lymphocytes from splenic, lymph nodes, bone marrow, and blood (but not thymic lymphocytes) has been shown (Ratt et al., 1970; Rabellino et al., 1970). Since all proteins and other molecular constituents of cells except deoxyribonucleic acid (DNA) are in a steady state of molecular turnover, it follows that these cells must be synthesizing antibody albeit very slowly. Therefore, it is only necessary to amplify the rate of synthesis of antibody at the cell level. To do so this cell must synthesize more protein for which more mitochondria, m-RNA, transfer (t)-RNA and ribosomes are needed. Their rate of energy exchange becomes greater. The cells hypertrophy (differentiate and transform) which in accordance with the "critical mass" concept[1] (Mazia, 1956, Johnson, 1969) leads to mitosis and amplification by increased cell numbers.

In terms of the preceding notions one can visualize the primary antibody response as follows. A few cells with little antibody trap a little antigen and their protein machinery is stimulated a little. Although there is short term proliferation of cells the process does not continue so as to yield a steep exponential increase. Only a low titer of antibody results. While the preceding continues, antigen is being degraded. Such a mechanism would appear to have a threshold, such that tiny amounts of antigen would not produce a detectible effect.

When the second antigenic stimulus is applied the amount of antibody per cell and the number of cells are greater. More antigen is trapped, and thus, the stimulus for antibody

[1] This concept says cells divide when they become too large and inefficient machines and implies that regulation of cell division is controlled by an internal cell feedback loop and not through feedback loops external to the cell.

synthesis and the concomitant proliferation is greater The stimulated cells expand exponentially now on the steep part of the curve Germinal centers "blow up" Exponential increases in antibody titer are observed with a steeper slope A natural concomitant of such a system must be a built in negative feedback mechanism that somehow turns off the multiplication processes A large number of cells do migrate into the gut, as will be discussed later

The modified instructive theory described above predicts that cells could respond to separate antigens and produce more than one antibody, whereas, the clonal selection theory predicts that the antibody produced by each clone should have a specificity for its antigen Some cells have been found producing two different kinds of antibody It appears, therefore, that neither the clonal selection theory nor the modified instructive concept is exclusively correct, and that some combination of the two will prove to be needed

IMMUNOLOGICAL UNRESPONSIVENESS, PARALYSIS, TOLERANCE

Although antigen is accepted as one of the prime regulators of antibody synthesis, the usual response of transformation and proliferation of cells producing the specific antibody does not always occur A second type of response is possible in which no immunological reaction is observed, at least for a period of time after introduction of the antigen there is no antibody production This *immunological paralysis* is achieved only under special circumstances by the use of very large doses of antigen (Felton, 1949)

In newborn animals and the fetus the failure to respond has been observed following the administration of many different kinds of antigen, including homologous cells and foreign proteins This is called *actively acquired immunological tolerance* (Billingham et al , 1954)

A potentially clinically important aspect is the induction of tolerance to tissues with histocompatibility differences If tolerance is induced in a newborn mouse by the injection of spleen cells from a different strain of animal, the injected animal becomes tolerant When it becomes an adult it will accept skin grafts from animals of the strain from which the cells were originally obtained The tolerance to these skin grafts is specific, because the recipient animal rejects grafts obtained from mice of a third, unrelated, strain A thorough understanding of tolerance might lead to mechanisms by which transplantation of organs could be accomplished with impunity

THE THYMUS IN IMMUNITY

Current thinking relates the thymus to the development of the cellular immunity shown in allograft rejection and antibody formation via the concept that its cells are derived from the mammalian equivalent of the lymphoid bursa in birds (Meuwissen et al , 1969) Although this is an attractive hypothesis there is no clearcut evidence for the existence in mammals of an equivalent of the bursa of Fabricius in birds For the time being an open mind on the subject is required

Thymectomy of the newborn mouse originally led to the notion that the thymus is essential for the development of immune responses, since in initial studies the mouse thymectomized at birth accepted skin grafts from unrelated mice and often failed to respond to antigen with the production of circulating antibody (Miller et al, 1962) However, in repetition of these studies on pathogen-free animals (Hess et al , 1963) or the germ-free animal (Miller 1967), removal of the thymus did not suppress humoral or cellular immunity, and the animals responded normally to antigens and rejected the skin grafts (Hess et al , 1963) No one questions the fact that the thymus is related to immunity, but its precise role is not yet clearly established Thymectomy in the adult diminishes the lymphocyte count in the blood and output of thoracic duct lymphocytes, but has little influence on immune process Perhaps the sole function of the thymus is the production of vast numbers of lymphocytes

Lymphocytes

The lymphocytes occupy a central position in humoral and cellular immunity and as such their characteristics and behavior are essential to an understanding of these phenomena

HETEROGENEITY OF LYMPHOCYTES

The lymphocytes constitute a family of cells of different tissue origins, migration patterns, size, staining characteristics, ultrastructure, lifespan, and function Thoracic duct lymphocytes consist of two overlapping populations on the basis of volume (Sipe et al , 1966) The

smaller cell has a median volume of 270 μm³ and the larger median volume of 645 μm³. Morphologically the *small lymphocytes* are relatively featureless, consisting of a compact nucleus with a nucleolus and cytoplasm containing a Golgi apparatus, a few mitochondria, ribosomes and a poorly developed endoplasmic reticulum. The small lymphocytes appear to be metabolically relatively inactive. The *larger lymphocytes* have more numerous ribosomes and in some instances have polyribosomes with a rapidly developing endoplasmic reticulum. There are all gradations between these cells and mature *plasma cells* that contain an exceptionally well-organized endoplasmic reticulum to which are attached innumerable ribosomes (Fig. 4-34). The plasma cells thus possess the full efficient apparatus for protein synthesis. All stages of lymphocytes ranging from the small featureless cell to the characteristic plasma cell synthesize protein and specific antibodies.

The lifespan of lymphocytes varies from a few hours to years. By continuous infusion of tritiated thymidine into animals to label each lymphocyte as it is born, it has been shown in rats that the class of smaller lymphocytes has a median survival time of about 1 month while 5 to 8% had lifespans greater than 9 months (Robinson et al., 1965). The larger lymphocytes also are divided into cells with lifespans of the order of 1 to 2 days and others with a maximum lifespan of about 60 days (Little et al., 1962). It has been suggested that lymphocytes in human females have a mean lifespan of 530 ± 64 days (Norman et al., 1965). The long lived lymphocytes are believed to be immunologically committed and to serve as a basis for immunological memory (Fitzgerald 1964, Gowans and Uhr, 1966). The long lived cells are produced at a rate proportional to body growth, and recirculate from the blood to lymph (Everett et al., 1964). Heterogeneity of function will be discussed later.

THE GENERATIVE CYCLE OF LYMPHOCYTES

Methods described in Chapter 2, using tritiated thymidine labeling of DNA and observation of the flow of cells through mitosis, have revealed many parameters of the cell cycle of lymphocytes. In the rat spleen, the lymphatic nodules exhibit a DNA synthesis time of 5 hr, a minimum [illegible] hr, with a maximum of 2 hr, and a generation time of 10 hr. It appears that almost [illegible] of the germinal center cells are in cycle although only 1 to 2% of the cells in the mantle zone surrounding the germinal center are in cycle. About 45% of the mantle cells probably are in G₀ state (Fliedner et al., 1964c). Using lymphocytes from the thoracic duct of rodents a systematic study has shown that the generation time is about 5 to 6 hr, with a DNA synthesis time of 4 hr and a mitotic time of about 15 min (Sainer et al., 1967). Lymphocytes have the shortest cell cycle of mammalian cells.

During antigenic stimulus the time that it takes for a population of cells to double in number is about 24 hr (Cabalbo and Makinodan, 1964). Since this doubling time is substantially longer, than the generation time described above it appears that there is some death of lymphocytes during antigenic stimulation since the doubling time and the generation time are equal only if all of the labeled cells divide and there is no death of the progeny.

Studies on lymphocyte production shows that 0.5 to 2.7% of the DNA of rodent lymphoid tissue is renewed per hour in lymph nodes and the thymus (Schooley, 1964). It is easy to measure by DNA labeling the number of cells emerging from the thoracic duct and the fraction of these that are newly produced. It is still very difficult, however, to get any realistic estimate of the absolute production rate of new lymphocytes in all of the different lymphocytic organs. There are too many uncertainties in determining the migration pattern, the fraction of cells that are recycling from blood to lymph, the reutilization of DNA labeling materials, and the possibility of death of cells from antigenic or other stimuli.

Under normal conditions the mass of the adult lymphoreticular system is relatively constant, implying that there is a balance between the production and death of cells produced in response to some small constant input of antigens absorbed from the gastrointestinal tract, from the skin or elsewhere. One can determine the number of newly produced lymphocytes that leave a single lymph node by combining efferent duct cannulation with arterial perfusion using tritiated thymidine in order to label the DNA of all new cells. The absolute production can be determined from the replacement

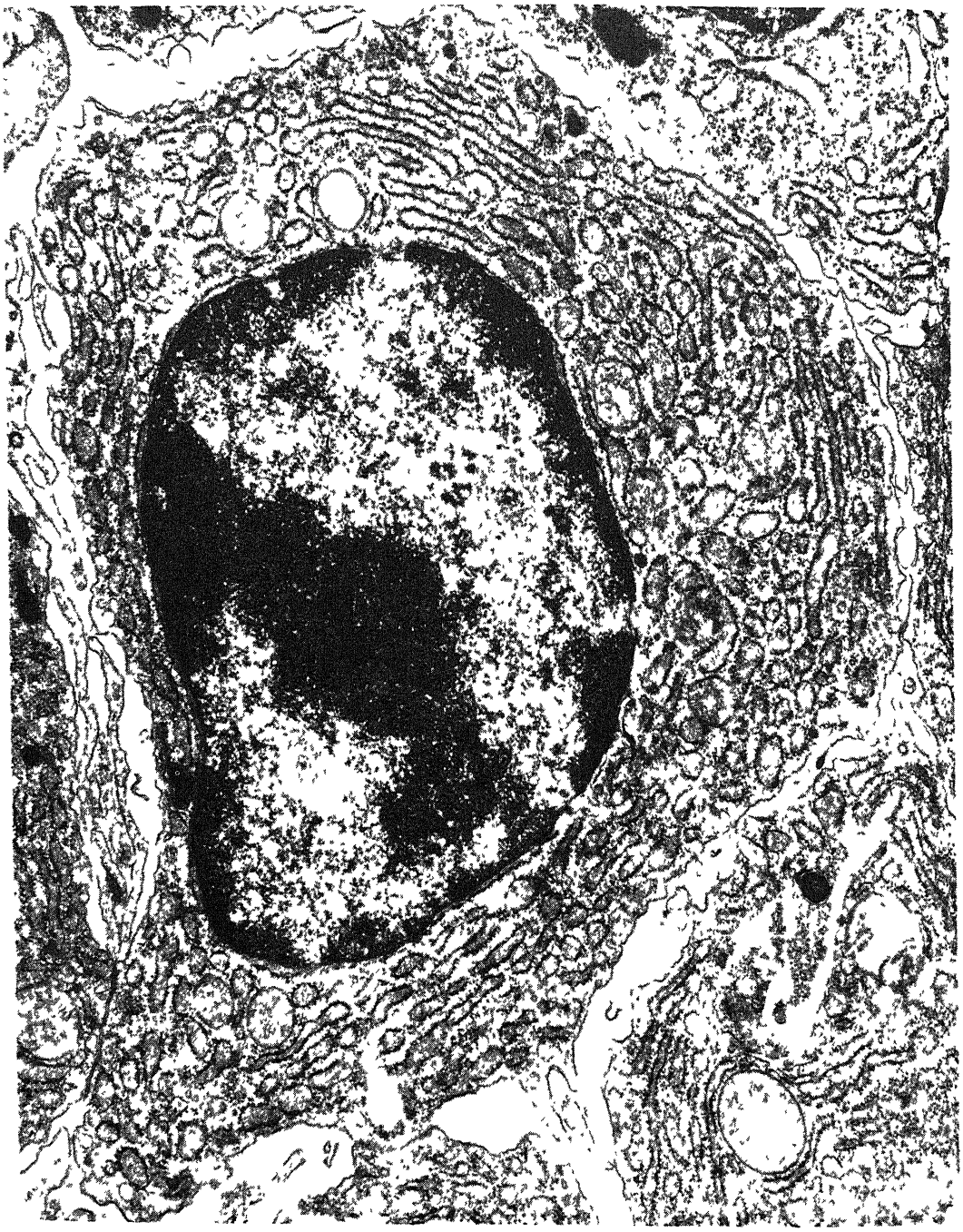


Fig 4.34 This is an electronmicrograph of a human plasma cell obtained by marrow aspiration The cytoplasm contains a well developed endoplasmic reticulum to which are attached ribosomes Magnification × 24 000 (Picture courtesy of Dr A F Howatson Preparation by Miss M Nagai)

of unlabeled by labeled DNA per unit time divided by the average DNA per cell, plus the number of labeled cells leaving the node The replacement of unlabeled by labeled DNA in the steady state represents the intranodal cellular death rate of cells produced in the node Imperfect knowledge of lymphopoiesis, however, prevents an accurate assessment of the

absolute production rates that can be determined for erythropoiesis and granulopoiesis

RECIRCULATION OF LYMPHOCYTES AND MIGRATION PATHWAYS

Examination of lymph from the thoracic duct in several mammalian species has shown that the output of lymphocytes is sufficient to replace all of those in the circulating blood several times daily (Drinker and Yoffey, 1941) The actual number of cells that enter the blood from the thoracic duct each day is very large, of the order of 10^9 in the rat and proportionately more in larger animals In the steady state whatever number of cells enters the blood must either be destroyed, migrate elsewhere, or recycle from the blood to the lymph and back again Two techniques have made possible the solution of this problem First, a technique for collecting lymph from the thoracic duct of unanesthetized rats over long periods was developed (Bollman et al, 1948) Second, tritiated thymidine and other pyrimidines were introduced into the study of cell kinetics in mammals (Hughes et al, 1958) and into the study of lymphopoietic tissue (Cronkite et al, 1959) The combination of these two procedures provided the answer that most of the turnover of lymphocytes is more apparent than real, since the same lymphocytes continuously recycle from blood to lymph through the postcapillary venule[2] of the lymph node, and then are swept back into the lymph as it filters through the sinuses of lymph nodes (Gowans, 1957, 1959, Everett et al, 1964, Gowans and Knight, 1964) The route of the migration of the lymphocytes through the postcapillary venules is believed to be through the high endothelial cells (Marchesi and Gowans, 1964) However, lymphocytes also proceed between endothelial cells (Smith et al, 1970), like granulocytes and monocytes Whether this endothelial transit is unique can be questioned since red cells also proceed through the endothelial cell under some circumstances (Johnson et al, 1966) In addition endothelial cells are phagocytic at times and actively engaged in pinocytosis

Much is now known about the migration pathways of the ubiquitous lymphocyte (Fig 4 35) The old concept that there is a major migration stream of small lymphocytes from the blood to the marrow, accounting for the daily replacement factor (DRF), and that these migrants are transformed and serve as stem cells (Yoffey and Courtice 1956) is not now accepted On the contrary, there is no doubt about a substantial proliferation of lymphocytes within the bone marrow and their migration into the blood (Osmond and Everett, 1964 Everett and Caffrey, 1967) The evidence for proliferation of small lymphocytes in the marrow is unequivocal even though the nature of the precursor cell is not clear It is conceivable that the small lymphocyte is in a long G_1 or G_0 phase and undergoes transformation and proliferation The large scale intramedullary production of lymphocytes as well as the minor migration of lymphocytes into the marrow demands either a large scale intramedullary death or else a migration of cells out of the marrow to maintain steady-state conditions Lymphocytes have been seen migrating between the sinusoidal endothelial cells of the marrow (Hudson and Yoffey, 1966) Elegant studies on marrow transplantation using chromosome markers have shown that the bone marrow contains cells capable of repopulating the aplastic thymus, spleen, lymph nodes, and the bone marrow of otherwise fatally irradiated mice (Ford et al, 1956) This depends upon a slow migration of cells that takes several weeks to achieve complete regeneration of hematopoietic tissues The migration rate of marrow lymphocytes under steady-state conditions in the adult has not been established There is also some evidence that bone marrow derived cells are the precursors of antibody-synthesizing cells (Miller, 1967, Mitchell and Miller, 1968, McGregor, 1968)

The evidence for migration of cells out of the thymus during fetal life and in the newborn mouse is likewise unequivocal (Michalke et al, 1969) The lymphocytes leave the thymus both by lymphatics and by blood vessels (Kotani et al, 1966, Iorio et al, 1970) They migrate as shown in Figure 4 36, becoming concentrated first in the spleen at the marginal zone as illustrated in Figure 4 17 of Chapter 3 Apparently the lymphocytes flow at the periphery of the blood in the central arteriole of the white pulp Plasma rich in lymphocytes is skimmed out into the small branches penetrating to the marginal zone (Weiss, 1962), where the labeled

[2] An anatomically characteristic venule with cuboidal endothelium located in paracortex of lymph nodes and seen in tonsils and Peyer's patch Lymphocytes migrate through the wall of this vessel

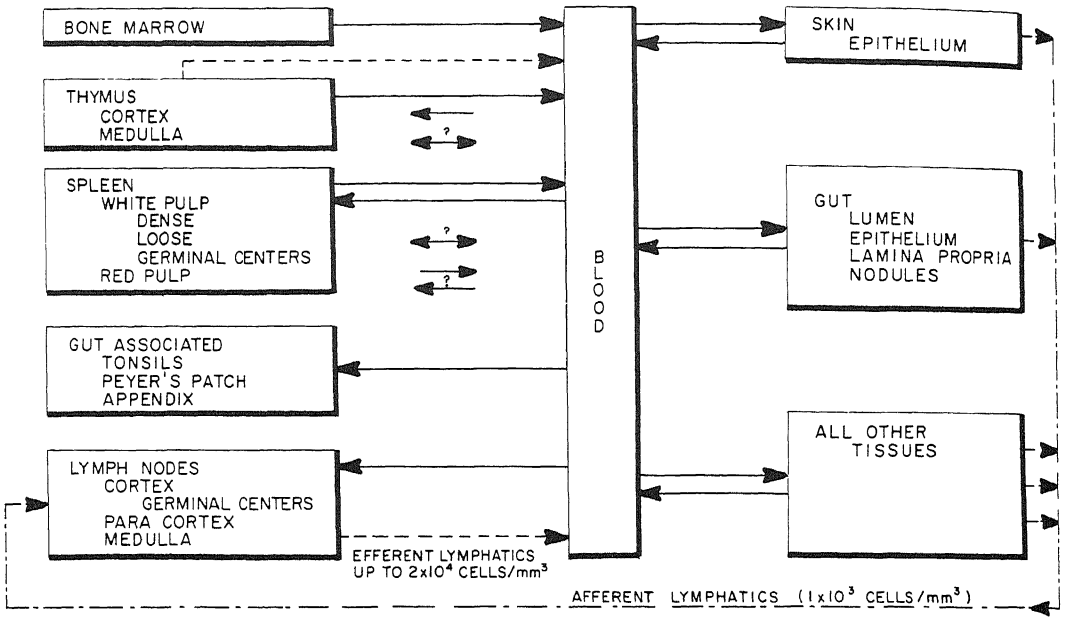


Fig 4 35 The ubiquitous lymphocyte and its exchanges with blood and tissues

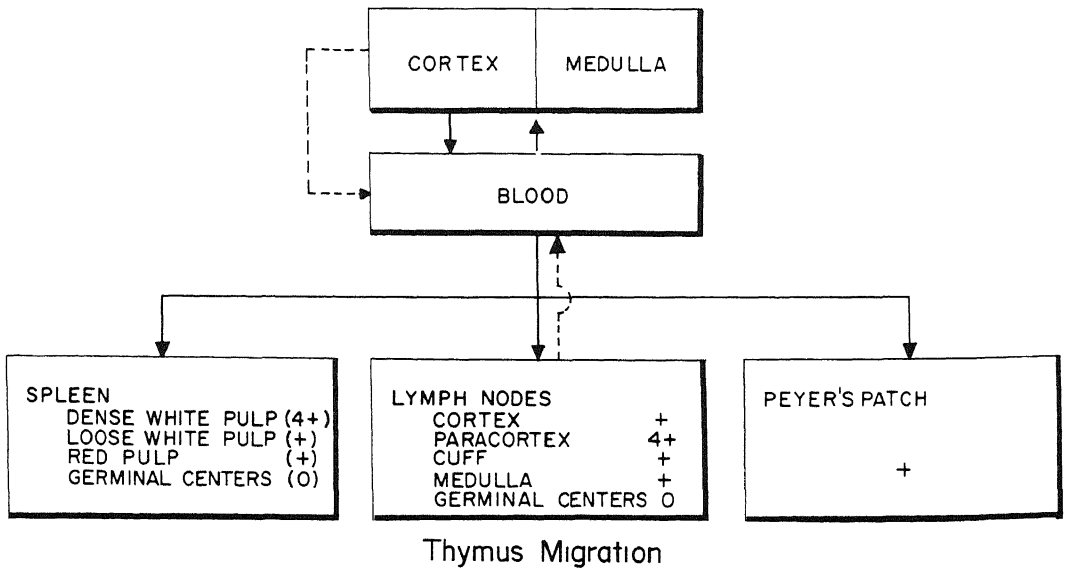


Fig 4 36 Schematic routes of migration of thymic cells showing sites of relative concentrations in the tissues and the return to the corticomedullary junction

lymphocytes are first seen From there they migrate into the periarteriolar lymphoid sheath, then move towards and concentrate around the central arteriole of the white pulp (Goldschneider and McGregor, 1968, Ford, 1969) from which they enter the blood by an undisclosed route The spleen contains no blood vessels similar to the postcapillary venules of lymph nodes

The migration of lymphocytes out of the blood is not exclusively through the endothelium of the postcapillary venules since many cells are also found moving in between the endothelial cells After the cells leave the ven-

ules they pass through gaps in the walls of the lymphatic capillaries and are swept off in the lymph stream The mechanism that directs this large scale, recycling traffic through the endothelium of the postcapillary venules is not understood It has been shown, however that treatment of lymphocytes with trypsin and neuramidase prevents the entry of lymphocytes into the lymph nodes but does not influence their migration into the spleen (Woodruff and Gesner, 1968) These observations suggest that sialic acid plays an essential role in allowing the cells to migrate through the postcapillary endothelium

Although the paracortical areas of the lymph node and the dense white pulp of the spleen have been defined as thymus dependent areas (Parrott et al, 1966), this concept can be questioned Long term studies on the migration of labeled thymic lymphocytes show that these cells "home" not only on the so-called thymic dependent areas but also migrate out into the red pulp, the outer cortex of the lymph nodes and into the medulla An occasional thymic cell migrates back to the thymus in the region of the corticomedullary junction These cells appear to be short lived or else divide rapidly, since none was found in the tissues 96 hr after termination of the study of Iorio et al, (1970)

There is a considerable production of lymphocytes in the spleen Moreover, lymphocytes recycle through the spleen at a rate proportional to their concentration in blood, with a minimum transit time through the spleen of about 2 to 3 hr The route of migration involves the trapping of lymphocytes in the marginal zone from which they enter the white pulp Labeled lymphocytes are first seen in the marginal zone, later in the dense white pulp of the periarteriolar sheath, and then concentrated around the central arteriole and later in the red pulp The lymphocytes that recycle through the spleen can also recycle through the lymph nodes There is also a well-established recycling of lymphocytes through postcapillary venules in the Peyer's patches of the gut A substantial production of lymphocytes occurs in the gut-associated lymphoid tissue with some newly formed cells entering the blood in addition to the recycling There is also a continual migration of lymphocytes into the lumen of the gut (Kotani, et al, 1967) By cannulating afferent lymphatics one can determine that there is a continual flow of lymphocytes from the blood through the interstitial tissue and back into the lymph nodes although the concentration of lymphocytes in normal afferent lymph is very low compared to efferent lymph

If the efferent lymphatic of a single lymph node is cannulated and drained continually the anatomic structure of the lymph node remains the same (Hall and Morris, 1962) These results indicate that the efferent small lymphocytes are recycling through the node from elsewhere in the body and that the large cells produced within the node continue to be produced at the same rate despite their loss from the body

Considerable information has been obtained on the size of lymphocyte pools and the kinetics of their depletion by prolonged drainage of lymph from the body or by extracorporeal irradiation of the blood or lymph In the latter technique either blood or lymph is diverted through a radiation source that delivers a dose sufficiently high to kill the lymphocytes Following their return to the circulatory system they are removed from the circulation and destroyed During depletion studies the thymic cortex is diminished in thickness, and the corticomedullary ratio is decreased Splenic lymph follicles are markedly decreased in size as a result of a loss of small lymphocytes from the cuffs surrounding the germinal centers and there is also a remarkable loss of lymphocytes from the less densely populated areas of white pulp, as well as from the red pulp The germinal centers in the spleen and lymph nodes are intact There is a striking diminution in the population of the paracortical areas of the lymph node and relatively little change in the outer cortex From these studies it has been concluded that lymphocytes in the outer cortex of the lymph node are relatively tightly fixed and mobilized slowly, whereas, the lymphocytes of the paracortical area are mobilized easily (Cottier et al, 1964b, Ruchti et al, 1970) In the rat the mobilizable lymphocyte pool is approximately 7.8×10^9 cells per kilograms of body weight (Caffrey et al, 1962) In the calf the lymphocyte depletion process consists of two exponential components The first component represents a pool of readily accessible lymphocytes that are mobilized with a half time of about 1.2 days The average size of this compartment is about 4.8×10^9 small lympho-

cytes per kilogram (Schnappauf and Schnappauf, 1968, Cronkite et al, 1968) The half time of depletion for the less accessible lymphocyte pool is of the order of 30 days Repletion of the lymphocyte pools follow also a two-component pattern One is relatively rapid and takes place in a few days, whereas, in some animals it has taken over a year for blood lymphocytes fully to attain predepletion levels A schematic presentation of the flow rate of lymphocytes from blood to tissues to lymph nodes based on studies in the calf is shown in Figure 4 37

The injection of complete Freund's adjuvant[3] subcutaneously produces a granuloma having blood vessels resembling postcapillary venules, and sustaining a flow of small lymphocytes in the draining lymphatic about equivalent to that of a lymph node of the same size (Morris, 1968) Regional antigenic stimulation of a lymph node induces transformation and proliferation of lymphocytes that can be seen flowing into the efferent lymph of the stimulated node These cells are secreting antibody and have a well-developed endoplasmic reticulum (Hall et al, 1967) The presumption is that these cells migrate into the tissues and complete differentiation into plasma cells If these lymphocytes are diverted from the body, the immune response is limited to the lymph node But, if they are returned to blood the immune response becomes generalized, thus, proving that the obligatory exit from lymph nodes is through the efferent lymphatics

Even though the cells that emerge from the lymph nodes via the efferent lymphatics after a local antigenic stimulus are antibody synthesizing cells, there remains some controversy as to their origin A concept now in vogue states that they originate in bone marrow, and come to reside in the outer cortex of lymph nodes There is an antigen-reactive cell believed to be thymic in origin (Mitchell and Miller, 1968)

CELLULAR IMMUNITY (DELAYED TYPE HYPERSENSITIVITY, ALLOGRAFT REJECTION AND IMMUNE SURVEILLANCE)

Immune responses can be transferred between genetically identical animals by the transfer of splenic, lymph node, blood, or thoracic duct lymphocytes from the sensitized animal to the nonsensitized animal The pioneering discovery of this phenomenon arose from the classic studies of Landsteiner and Chase's transfer (1942) of delayed cellular hypersensitivity by transfer of cells from the guinea pig This simple observation clearly separated cellular from serum antibody responses and provided an immunological test for analysis of the mechanisms of cellular immunity Living cells were required for the transfer, and the transfer was short lived between unrelated animals because the living cells were rejected In human beings it was possible to transfer delayed cutaneous hypersensitivity by a dialyzable extract of sensitized lymphocytes called the *transfer factor* (Lawrence, 1955) The transfer factor is specific for the antigen that originally sensitized the cells The factor is dialyzable, has a molecular weight of less than 10,000, and is nonantigenic It appears to be a polypeptide-polynucleotide, heat labile at 56° C It converts normal lymphocytes in vitro and in vivo to a specific antigen-responsive state These cells when confronted with antigen transform and proliferate (Lawrence and Valentine, 1970) In addition to delayed cutaneous hypersensitivity transfer factor can transfer allograft sensitivity

When transfer factor is added to lymphocytes in the presence of antigen another material is liberated called *lymphocyte transforming factor* (LTF) that is not dialyzable and distinct from histocompatibility antigens It functions as a "recruiting agent" for nonsensitive lymphocytes Ordinarily, only a very small fraction of lymphocytes become sensitized to a specific antigen When this antigen is reintroduced, there is a transformation and proliferation of the sensitive cells with an increase in the number of cells It is conceivable that the transfer factor and lymphocyte transforming factor are released in vivo and recruit additional nonsensitive cells that can then be engaged by the antigen and so further amplify or escalate the intensity of the immunological response

MIGRATION INHIBITORY FACTOR (MIF)

This factor inhibits the migration of macrophages It was discovered by Bloom and Bennett (1966) and David (1966) The material is released from sensitized lymphocytes by the specific antigen that produced the hypersensi-

[3] Freund's adjuvant is a mixture of mineral oil and an emulsifying agent, it is used as a vehicle for suspension of antigens to be injected subcutaneously

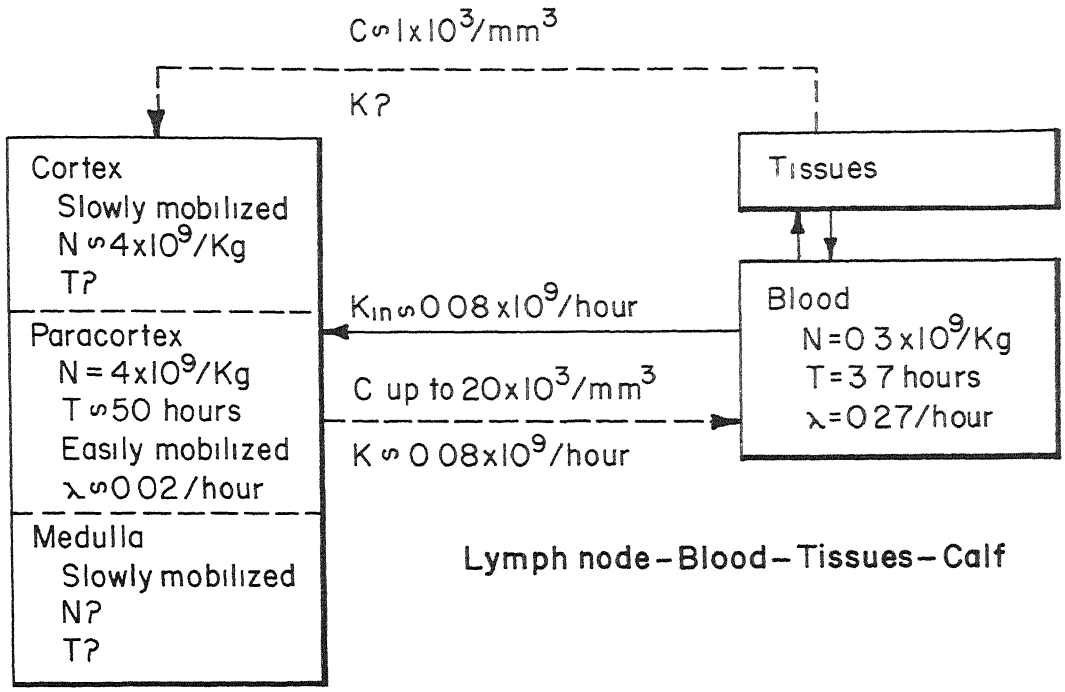


Fig 4 37 Schematic presentation of the flow of lymphocytes from blood to tissues to lymph nodes with estimates of time fluxes and sizes where data are available from experiments on calves N number/kg body T average time in compartment λ fractional turnover/hr K flux (number/hr/kg) C concentration numbers/mm³

tivity reaction Thus, it serves to retain macrophages at area of immunological action where they can attack and destroy the offensive material The detailed properties of MIF are reviewed by Bloom and Jimenez (1970)

LYMPHOTOXIN

This material is released from the lymphocytes following an interaction with the target cell on antigen, or after treatment of lymphocytes with agents inducing transformation such as phytohemaglutinin (PHA) It is difficult to study this reaction in vivo It can be envisioned, however, that whenever a lymphocyte carrying antibody is trapped by the specific antigen this lymphocyte may die and/or release its load of lymphotoxin, and thus, intensify local necrosis This substance may be a major effector in delayed sensitivity and allograft rejection (Granger, 1970)

The Immunological Response, Its Initiation, Perpetuation, and Suppression

The probable sequence of events in immune responses is shown in Figure 4 38 in a diagram that begins at the *allograft* Antigen seems to be transported from the allograft to the regional lymph node either by cells, possibly macrophages, or in solution through the afferent lymphatics Within the lymph node the antigen reactive and antibody synthesizing cells interact in a manner not well understood Cells are triggered into cell cycle from a G_0 stage, undergo hypertrophy as described earlier, and ultimately enter the proliferative cycle and divide In the process of repetitive divisions their protein synthesizing machinery continues to develop and they take on characteristic appearance of intensely basophilic cells with varying degrees of endoplasmic reticulum After 3 to 5 days of proliferation, cells commence to make their exit through the efferent lymphatic and proceed into the blood where they become metastasized through the lymphoreticular tissues, where they set up secondary foci if engaged again by new antigen These cells entering the blood have a large complement of cell bound antibody, and some cytotoxic antibody appears in the serum It can be supposed that as lymphocytes proceed through the allograft that antigen is recognized, and cells are directed out of the blood

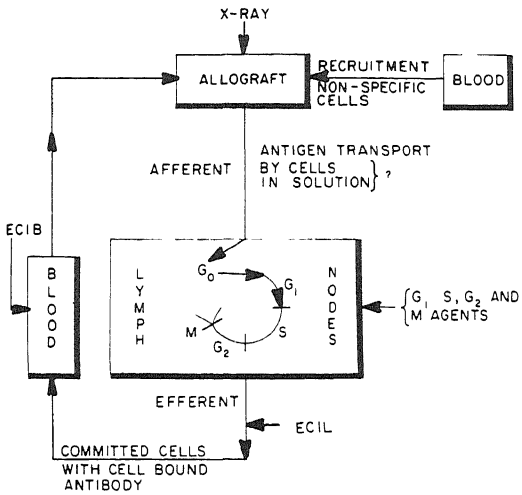


Fig 4 38 Schematic presentation of the role of lymphopoietic proliferation and migration in allograft rejection indicating the sites of attack by radiation and chemical to attain immunosuppression

stream into the allograft Following the entry of even a very few specifically immune lymphocytes, a series of events is initiated First, the cells release transfer factor and lymphocyte transforming factor, which as discussed earlier are believed to be capable of recruiting nonsensitized lymphocytes Upon contact with a specific antigen the cells also release lymphotoxin and migration inhibitory factor The lymphotoxin will set up an inflammatory reaction through death of the cells, which will further attract by chemotactic influences granulocytes and macrophages into the allograft The macophages will be immobilized by the migration inhibitory factor The two phagocytic cell types probably commence cleaning up the necrosis by phagocytosis of debris, and also escalate the inflammatory response by release of their lysosomal enzymes when they in turn are destroyed This outline of what might be happening is different from classical accounts of the reactions, in that here the cell bound antibody is relegated to a position of recognition and not primary destruction A similar sequence of events can be visualized as occurring during cutaneous delayed hypersensitivity if one considers the sensitized cells to be antigenically abnormal, and thus, requiring identification and destruction.

The concept of immunological surveillance of antigenically altered somatic cells has been extended to cancer It is believed, although not conclusively proved, that somatic mutations taking place at random result in an antigenically different cell that can be identified Consequently, the entire sequence of events described above is set into motion to remove this neoplastic cell When the individual either becomes tolerant or when immunological reactivity is suppressed for one reason or another, the tumors grow Whether this hypothesis is valid is not yet known, but it seems plausible and may lead the way to more effective cancer therapy

The notion that nonsensitive lymphocytes are recruited into a specific immunological reaction is enhanced by the studies of Pedersen and Morris (1970) who have shown that the entire immunological sequence of kidney rejection can take place inside of the graft without the help of the regional lymph node They did this by cannulating the lymphatics from a kidney and collecting all of the transformed cells in the lymph

By referring to Figure 4 38, one can also identify the means by which one can launch an attack against the immune response Starting at the lymph node one sees that chemical agents directed at any subdivision of the mitotic cycle may act by destroying cells in that phase For example, as cells emerge from a lymph node via the efferent lymph they can be irradiated to kill the cells, when anatomical relations are such as to make this possible (Chanana et al , 1969) Similarly, the blood can be irradiated to kill that portion of the cells that go through an extracorporeal irradiation shunt One can also irradiate the graft, killing the lymphocytes and other cells that have been recruited into the graft Nevertheless, the relatively poor success of these types of immunosuppression stems from the fact that in suppressing transformation and proliferation of the immunologically committed cells one also produces injury that may be lethal to the bone marrow and epithelial structures

HISTOCOMPATIBILITY ANTIGENS

These antigens (HL-A) are responsible for stimulating an immune response against a transplanted organ or tissue Because the histocompatibility alloantigens reside on cell

membranes they are becoming useful in the study of the surfaces of cells. In the mouse these antigens are glycoproteins. Progress is being made in determination of the amino acid composition, secondary and tertiary structure. Details of structure, experimental approaches and use of histocompatibility antigens has been the subject of a recent symposium (Amos 1970).

CELL MEDIATED IMMUNITY

This is a form of nonhumoral immunity involving *macrophages*. When the macrophage comes into contact with the specific antigen such as tubercle bacillus it becomes activated. The activated macrophage is larger and more complex morphologically and when transferred to glass, it spreads. It also has very active phagocytosis and an increased content of acid hydrolases, and actively proliferates. Its precursor is the blood monocyte. Although this type of cellular immunity persists for the duration of an infection, it is relatively short lived after the infection is over. The immunity is both specific and nonspecific in that when induced by one organism the macrophages throughout the body develop increased activity towards other unrelated organisms. As animals recover from infection their nonspecific resistance does disappear, but they remain resistant to the original infection and express the immunity in part, for example, by the tuberculin skin test, a delayed hypersensitivity. Living lymphocytes from convalescent donors are able to transfer this type of cell mediated immunity. It is exceptionally important in combatting tuberculosis, brucellosis, listeriosis, and other infections. Cell mediated immunity was reviewed by Mackaness (1970).

IMMUNOGLOBULINS

A common characteristic of all types of immunoglobulin is that they may carry antibody activity and that all have the same general 4-peptide chain structure. Two central "heavy" chains are connected by a disulfide bond, and a "light" chain is connected to each of the heavy chains by a disulfide bond. The light and the heavy refer to molecular weights. The immunoglobulins have a bewildering number of synonyms, and are now classified into IgG (*γ globulin*), IgM (*macroglobulin*), and IgA. In human serum the relative amounts are 71%, 7%, and 22%, respectively. Antigenically

[illegible] called. The approximate molecular weights are IgG [illegible], IgM [illegible], and IgA [illegible]. However IgA [illegible] and may have a considerable [illegible] molecular weight [illegible] the [illegible] IgA contain [illegible] IgM [illegible] and IgA [illegible]. The [illegible] and more complex [illegible] rapidly exceed the [illegible] the scope of this discussion.

There are genetically determined variants of particular subclasses of heavy chain or light chain that occur among members of a species. These markers are called allotypes and are useful in genetic studies in man and animals.

IgG is often designated as 7S γ-globulin with reference to the Svedberg sedimentation units, the IgM type as 19S γ-globulin. The light chains from the immunoglobulins of any one individual have the same amino acid composition, molecular weight, electrophoretic mobility, and antigenic specificity. The heavy chains of the three classes are distinct and different in all four of the preceding properties. For details of structure in relation to function see reviews by Deutsch and Fudenberg (1969), Franklin and Frangione (1969), and Cebra (1969).

A newborn infant is relatively speaking immunologically undeveloped, susceptible to bacterial infections that are not pathogenic in older children or adults. Some of the antibodies in the plasma of the newborn infant are acquired by transfer from the maternal circulation. Thus the infant at birth is relatively well endowed with IgG from its mother. Very little IgM or IgA is transferred from the mother. IgA is closely related to defense against intestinal organisms, and the deficiency of this fraction explains in part why the newborn infant is susceptible to Gram negative infections. Infants degrade the maternally transferred IgG relatively rapidly so that practically none exists by 10 weeks of age. In about the 4th to 5th week of age, infants commence synthesizing their own IgG and other immunoglobulins, and by 12 months of age have plasma concentration nearly equal to that of the adult.

Multiple myeloma is an abberation in which there is increased synthesis of abnormal immunoglobulins. Any of the immunoglobulin types may be involved. The Bence Jones protein found in the urine of multiple myeloma

patients has been shown to be the light chain. The bone marrow characteristically shows an infiltration with myeloma cells that resemble plasma cells of varying degrees of immaturity. There is a high concentration of γglobulin in the blood: the disease is fatal.

DISEASES DUE TO IMMUNOLOGICAL DEFICIENCIES

Deficiencies in antibody production are primary or acquired. The *physiological hypogammaglobulinemia* of the infant has already been discussed. Bruton's disease is a sex-linked familial hypogammaglobulinemia in which immunoglobulin levels of 50 mg. or less per 100 ml. of blood are common. The individuals are very susceptible to infection. Secondary hypogammaglobulinemia may develop in the course of chronic lymphocytic leukemia and other diseases. There is a general decrease in plasma cells in the bone marrow, lymph node, and other tissues. Lymph nodes from primary hypogammaglobulinemia show ill-defined germinal centers and a thin cortex; there is slight development of Peyer's patches in the gut; yet, the thymus appears to be relatively normal.

There is a group of diseases in which there is an impaired cellular immunity called *thymic alymphoplasia*. There is little if any immunoglobulin deficiency. There is a lymphopenia. The thymus is very poorly developed, weighing as little as 1 g. Thymus, spleen, lymph nodes, tonsils, and gastrointestinal tract have a markedly reduced content of lymphocytes. There is a marked susceptibility to infections, particularly of mycotic and virus origins.

In another disorder in which there is a combined deficiency of antibody production and cellular immunity, antibodies are practically absent, there is a marked lymphopenia, and the patients are so very susceptible to infection that they die shortly after birth. There are two additional types of thymic alymphoplasia based on inheritance. Both have impaired delayed hypersensitivity reactions, hypogammaglobulinemia, very poor antibody response, few plasma cells, and an atrophic lymphatic tissue.

REGULATION OF THE HYDROGEN ION CONCENTRATION OF BODY FLUIDS

Hydrogen Ion Balance

The normal concentration of H^+ ions in the extracellular body fluids ranging as it does from 44 to 36 nM/liter, and represented by a pH variation from 7 36 to 7 44 (Table 4 16), is the result of a continuously maintained dynamic equilibrium between the intake and production, and the suppression and excretion of hydrogen ions

SOURCES OF HYDROGEN ION

Although a small amount of hydrogen ions may be taken in with the food or as medication, most of them arise from the numerous metabolic activities which are going on unceasingly in the living individual The major sources of new hydrogen ions include

1) Oxidation of Carbon Compounds to Produce Carbon Dioxide. Carbon dioxide is the ultimate end product of the oxidation of carbon compounds In an average adult at rest, about 200 ml (almost 10 mM) of this gas is produced in the tissues, transported by the blood, and eliminated by the lungs per minute, or about 10,000 to 14,000 mM/day The rate of its formation depends upon the rate of metabolism and during severe exercise it may be increased to 10 or more times this amount Carbon dioxide is soluble in water and the actual concentration of CO_2 in the tissue fluids at any one time is determined by the partial pressure of the gas in these fluids This in turn is regulated by the respiratory system, which normally keeps the P_{CO_2} in the pulmonary alveoli and in the arterial blood at about 36 to 44 mm Hg At this partial pressure the concentration of CO_2 in the plasma is 1 1 to 1 3 mM/liter and this is always in equilibrium with a small amount of hydrated carbon dioxide or carbonic acid

$$CO_2 + H_2O \rightleftharpoons H_2CO_3 \rightleftharpoons H^+ + HCO_3^-$$

Any condition which prevents the lungs from excreting carbon dioxide as rapidly as it is formed, or causes them to excrete more than is being produced, will of course result in a change in the CO_2 and carbonic acid concentrations of the body fluids and this, if not counterbalanced, may cause significant changes in blood pH

2) Conversion of Neutral Food Stuffs to Organic Acids A great variety of moderately strong organic acids are produced during the normal intermediary metabolism of neutral carbohydrates, fats and proteins and equivalent amounts of H^+ ions are temporarily released into the body fluids Most of these intermediary acids are oxidized to CO_2 and H_2O or otherwise metabolized and in this way are eliminated almost as rapidly as they are formed As long as this occurs their hydrogen ions disappear again, and no effect is produced on the overall hydrogen ion balance At times, however, because of excessive metabolic activity, or because of the presence of certain diseases or for some other reason, these intermediary organic acids, especially lactic and pyruvic acids may be produced in greater quantities than can be metabolized or excreted When this occurs there will be a greater or less accumulation of H^+ ions and a fall in the pH of the body fluids

3) Oxidation of the Sulfur in Organic Compounds The oxidation of the sulfur containing amino acids methionine and cysteine yields equivalent amounts of hydrogen ion and sulfate ion This can be illustrated by the equation (greatly simplified)

$$2C_5H_{11}N_1O_2S + 15O_2 = 4H^+ + 2SO_4 + (NH_2)_2CO + 7H_2O + 9CO_2$$

methionine — urea

That H^+ are produced in this way is shown by the finding that the administration of a methionine load to an individual is followed by an increase in the renal excretion of H^+, which parallels very closely the excretion of sulfate ions (Relman, 1964)

4) Formation of Inorganic Phosphate. The hydrolysis of phosphoesters and the breakdown of phosphoproteins, nucleoproteins, and phosphatides in foods, result in the formation of inorganic phosphates and the liberation of corresponding amounts of H^+ (Relman, 1964) For the orthomonoester, this can be illustrated by the following equation

$$R{-}O{-}\underset{\displaystyle OH}{\overset{\displaystyle O}{\overset{|}{\underset{|}{P}}}}{-}OH + H_2O = R\,OH + \frac{0\,8HPO_4^=}{0\,2H_2PO_4^-} + 1\,8H^+$$

(Ratio of the phosphates at pH 7 4 is 4 1)

Table 4 16
Range of reaction of some body fluids

Fluid	pH	Hydrogen Ion Conc Nanomols per l
Pure water	7 0	100
Blood		
Normal mean	7 40	40
Normal range	7 36–7 44	44–46
Very severe acidosis	6 8	160
Very severe alkalosis	7 8	16
C S F normal range	7 36–7 44	44–36
Pure gastric juice, normal	1 0	100,000,000
Urine		
Normal average, about	6 0	1000
Maximum acidity	4 5	31600
Maximum alkalinity	8 0	10
Intracellular fluid (muscle) (DMO)[a]	6 8	160

[a] Determined indirectly by the distribution of 5,5-Dimethyl-2,4 oxazolidinedione
Waddell and Butler, 1959

In normal persons, on the usual protein intake, the last three of the above processes produce only about 50 to 90 mM of H^+ per day, but as the acids from which these are derived are nonvolatile or "fixed" they cannot be excreted by the lungs and their elimination must, therefore, be brought about by the action of the kidneys The production of H^+, particularly by the last two mechanisms, will be increased by any acceleration of protein breakdown and especially by the elevated rate of endogenous protein destruction which follows a severe tissue injury or other cause of increased catabolism When these injuries are accompanied by severe renal failure, the increased production of these strong acids, coupled with the impaired renal ability to excrete them may lead rapidly to a serious accumulation of H^+ ions, and to a sharp and dangerous fall in the pH of the body fluids

5) Medication Medication with ammonium salts or with mineral or organic acids, such as mandelic acid, may occasionally add significantly to the load of H^+ ions that must be handled by the body Ingestion of ammonium salts is equivalent, as far as supplying H^+ ion is concerned, to the ingestion of the corresponding amount of the free acid, since each ammonium ion taken in is converted in the liver to urea with the liberation of a proton

$$2NH_4Cl + CO_2 = 2H^+ + 2Cl^- + H_2O + (NH_2)_2CO$$

EXCRETION AND FIXATION OF H^+ IONS

The addition of new H^+ to the body, whether from intake or metabolism, is balanced by the action of chemical or physiological mechanisms which normally remove them as rapidly as they are formed so that the $[H^+]$ of the body fluids remains reasonably constant These mechanisms include

1) Dilution Hydrogen ions formed in excess at any one site—as for example in an exercising muscle—are rapidly distributed by the circulation throughout the whole of the extracellular fluid Although not a regulatory mechanism in the usual sense, the dilution brought about in this way will help to minimize the local increase of H^+ concentration which would otherwise occur

2) Buffer Systems A variety of buffer systems is present in the extracellular and cellular fluids of the blood and tissues These function as the first line of defense in preventing alterations in $[H^+]$ from being produced by the addition of acids or bases, and to help keep the H^+ concentration of these fluids within the normal range (Fig 4 39) These buffer systems are

A) In the Blood Plasma i) The bicarbonate/carbonic acid buffer system Quantitatively the most important buffer in the plasma is the one made up of sodium bicarbonate and carbonic acid The normal ratio of the two components of this buffer system in the plasma is

$$\frac{[HCO_3^-]}{0\ 0301 \times P_{CO_2}\ (43\ mm\ Hg)} =$$

$$\frac{[HCO_3^-]}{[H_2CO_3]} = \frac{26\ mM/liter}{1\ 3\ mM/liter} = \frac{20}{1}$$

As the pK′ of the bicarbonate-carbon dioxide system is 6 1, it can be calculated from the Henderson-Hasselbalch equation that as long as this

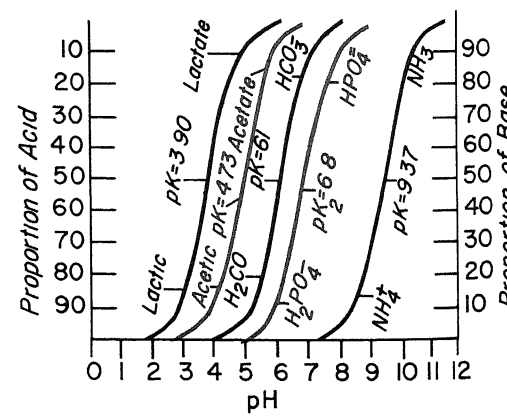


Fig 4 39 Calculated titration curves for several buffer systems showing the relation between [Base]/[Acid] ratios and pH For a discussion of pH titration curves and buffer mechanisms see Clark (1952)

ratio of bicarbonate and carbonic acid is maintained the pH will be 7.4

$$pH = pK' + \log [HCO_3^-]/[CO_2]$$

$$= 6.10 + \log 20/1.0 = 6.10 + 1.30 = 7.40$$

The very great importance of this buffer system in the control of the extracellular pH and its association with the respiratory regulation of carbon dioxide excretion will be discussed below.

ii) The phosphate buffer system. This consists of the mixture of the alkaline and acid phosphates of sodium

$$Na_2HPO_4/NaH_2PO_4$$

In this buffer pair the acid phosphate may be regarded as a weak acid (pK_{A_2}' for phosphoric acid is 6.8) and the alkaline phosphate as the fully ionized salt of the buffer pair. The normal ratio of base to acid in this case at the pH of normal blood is 4/1.

$$pH = pK_{A_2}' + \log ([base]/[acid])$$

$$= 6.8 + \log (4/1) = 6.8 + 0.60 = 7.4$$

iii) Plasma proteins. Plasma proteins act as buffers and, at the pH of the blood, they behave as a mixture of poorly dissociated weak acids and their almost completely ionized sodium salts

$$Na^+ \text{ proteinate}^-/H \text{ proteinate}$$

B) In the Red Blood Cells. Important buffer systems, distinct and separate from those of the plasma, are also present and active within the red blood cells and these are

i) Intracellular organic phosphate complexes such as adenosine diphosphate, creatine phosphate, etc.

ii) $\dfrac{\text{Potassium oxyhemoglobinate}}{\text{H oxyhemoglobin}}$ or $\dfrac{K^+ HbO_2^-}{HHbO_2}$

iii) $\dfrac{\text{Potassium hemoglobinate}}{\text{H hemoglobin}}$ or $\dfrac{K^+ Hb^-}{HHb}$

Within the red blood cells the oxygen-carrying protein, hemoglobin, acts as a buffer to H^+ ions much in the same way as do the plasma proteins, but in this case the buffering is done mainly by the imidazole groups of the histidine in the Hb molecule (Fig. 4.40). There is about four times more red cell hemoglobin than plasma proteins present in a given volume of whole blood (150 g. to 40 g./liter), the protein buffering capacity of whole blood, therefore, is about six times that of a similar volume of plasma. Both the oxygenated and the reduced Hb systems are depicted above, because oxidation of Hb to HbO_2 causes it to become a stronger acid, that is, to donate more H^+ ions, whereas, the deoxygenation of HbO_2 to Hb causes it to become less acid and to become more effective as an acceptor of H^+ ions. The formation of the more acid oxyhemoglobin in the alveoli by the oxidation of reduced hemoglobin will, therefore, compensate in part for the loss of H^+ ions and for the decrease in H^+ concentration at this

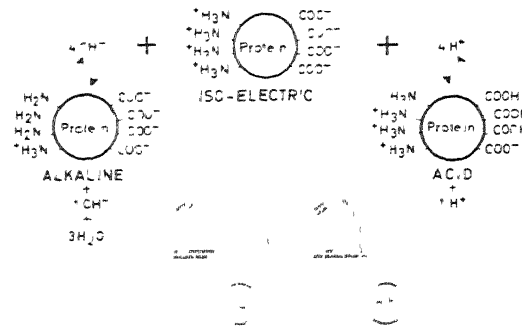


Fig. 4.40 Schematic representation of the buffering of acids and bases by plasma proteins and by the imidazole groups of hemoglobin histidine. (Adapted from Davenport, 1958.)

site resulting from the excretion of CO_2. In the same way the deoxygenation of arterial HbO_2 to Hb in the peripheral circulation where oxygen is given up, enables the hemoglobin molecule to accept and to fix more of the extra hydrogen ions added to the blood by the addition of carbon dioxide from the tissue cells.

C) In the Extracellular Fluids and Tissue Cells. Only a portion of the total buffer capacity of the body resides in the buffers of the blood plasma and red blood cells. The interstitial fluids have a bicarbonate and phosphate buffer concentration similar to that of the circulating plasma, and these play their part in all buffer reactions. In addition to these buffer systems, however, it has been shown by Pitts and his co-workers (1963) and by others, that somewhat more than 50% of the hydrogen ions infused as strong acid intravenously into nephrectomized animals, are transferred into tissue cells in exchange for sodium and potassium ions. These exchange reactions occur across all tissue cell membranes and make available in this way the large quantities of cellular protein and organic phosphate complexes for the buffering of extracellular hydrogen ions. The superficial surfaces of the bones may also play a role in these reactions. Similar ionic exchanges in the reverse direction allow the cellular constituents to take part in the buffering of extracellular base whenever this becomes necessary.

3) The Handling, Transport, and Excretion of Carbon Dioxide. This subject is fully covered in Section 6, Chapter 1 "Uptake and Delivery of Respiratory Gases" so that there is no need here to do more than to review it quickly in relation to H^+ regulation.

The large amounts of potentially acid-producing carbon dioxide, continuously being released from all the body tissues by cellular metabolism, are handled without seriously disturbing the $[H^+]$ of the body fluids in a manner which may be briefly described as follows.

The ultimate regulation of the pH level of the ar-

terial blood is carried out by the respiratory system, which, by controlling the minute volume of respired air, is able to vary the amount of CO_2 excreted to the outside air in such a way that the $[HCO_3^-]/H_2CO_3]$ ratio in the plasma is kept relatively close to the normal value of 10/1 no matter how much or how little carbon dioxide is being produced At this ratio, the pH will be 7 40 If the quantity of CO_2 being removed from the blood by respiration is not sufficient to keep up with that being added by metabolism, the P_{CO_2} and the carbonic acid content of the arterial plasma leaving the lungs and passing through the respiratory center will be slightly higher than normal The $[HCO_3^-]/[CO_2]$ ratio and the pH of this blood will be slightly depressed and the $[H^+]$ will be slightly elevated As a consequence the respiratory center, being very sensitive to changes in the pH and P_{CO_2} of the blood passing through it, will normally be stimulated to increase the degree of pulmonary ventilation More CO_2 will be exhaled and removed from the venous blood and this will continue until the $[HCO_3^-]/[CO_2]$ ratio and the pH of the blood leaving the lungs are restored to normal again On the other hand if too much CO_2 is being lost by respiration this ratio will be increased, and the pH of the arterial blood will be slightly elevated This change in the pH of the blood passing through the respiratory center will cause the pulmonary ventilation to be decreased Less CO_2 will be excreted and more will be retained in the arterial blood until the ratio and pH return once more to the normal values

It is not to be inferred from the above that the plasma bicarbonate buffer system and its regulation by the respiratory apparatus are alone responsible for the buffering of all the changes in $[H^+]$ which may be produced by the addition or loss of acids or bases to the body fluids All the buffer systems of the body are in equilibrium with one another and the effect of the respiratory control on the bicarbonate system is transmitted to all the others

4) Metabolic Events Such as the Oxidation or Utilization of Certain Organic Acids. Most of the organic acids, except uric acid, produced during the breakdown of normal food stuffs, appear only momentarily as intermediary products of metabolism and, as they usually disappear as quickly as they are formed, they normally have little effect on the hydrogen ion regulation Occasionally, however, as for example during severe exercise, lactic acid may be produced in amounts larger than can quickly be metabolized and, in spite of a large increase in respiration and the "blowing off" of large amounts of CO_2, the pH of the blood may be lowered appreciably below the normal range The subsequent disappearance of the lactic acid, resulting from its oxidation or conversion to glycogen, removes the extra protons associated with its accumulation, replenishes the depleted store of HCO_3^- ions and restores the pH back to normal Similarly the low levels of blood pH and of plasma bicarbonate which are present in severe diabetic acidosis are at least partially restored to normal when the excess acetoacetic acid present in the blood is destroyed following adequate treatment with insulin and glucose

5) Renal Mechanisms in the Regulation of Hydrogen Ion Concentration A very important part in the control of the $[H^+]$ of the body fluids and in the regulation of the overall H^+ ion balance is also played by the kidneys This particular activity of the kidneys, along with the many other aspects of renal function, is fully discussed in Section 5, Chapter 1 and little therefore needs to be said about it here In brief the kidneys are able, mainly through the activities of the cells lining the tubular lumina, to excrete widely varying quantities of H^+, both as titratable acid and as ammonium ions (NH_4^+), into the urine or to eliminate excessive quantities of base whenever either of these actions is necessary for the purpose of maintaining a normal level and balance of hydrogen ions The production and excretion of ammonium ions is dependent upon the enzyme, glutaminase I, contained in large amounts in the kidney Terminal renal failure is characterized by an inability of the kidney to continue to excrete excess hydrogen ions as ammonium ions and is typically evidenced by acidosis (An excellent review of the role of ammonium ions in acid-base balance has been published by Pitts, 1971) The tubular cells also help to maintain the concentration of bicarbonate ions in the blood and other extracellular fluids at the normal level of 24 to 28 mM/liter by varying the extent to which bicarbonate ion in the glomerular filtrate is reabsorbed back into the blood again or is passed on to the bladder urine to be excreted to the outside Also, when the normal store of bicarbonate ion in the body is depleted, as it may be by the addition of more nonvolatile acid than can normally be handled, the kidney tubules, in association with their excretion of H^+ ions, form large quantities of new bicarbonate ions, which, when added to the blood, help at least partially to restore the concentration of this ion towards its normal level The decrease in plasma HCO_3^- concentration resulting from diabetic acidosis, for example, would be much more severe and would occur much more rapidly than it does, if it were not for the ability of the renal tubules not only to excrete appreciable quantities of hydrogen ions but also, at the same time, to make new bicarbonate ions to take the place of those which have been used up in the neutralization of the excess acid

Disturbances in the Regulation of H^+ Ion Concentrations in the Blood

ACIDOSIS AND ALKALOSIS DEFINITIONS

In the healthy person the H^+ concentration of the blood and extracellular fluids is main-

tained, and the pH rarely varies more than 0.04 pH units on either side of the average pH 7.40. In a large number of pathological conditions the clinical and metabolic disturbances associated with, or caused by, the illness, subject the mechanisms responsible for the control of the pH of the body fluids to a considerable strain. In all such cases the ability of the body to resist changes in the pH of its fluids may at times be overcome and if this occurs the $[H^+]$ of the body fluids will, depending upon the nature of the underlying primary cause, be either increased or decreased to levels which are more or less beyond the normal range.

1) Acidosis[1] The blood and extracellular fluids may become more acid than normal. Here the pH is depressed to a greater or less degree below pH 7.36 and the hydrogen ion concentration will be correspondingly higher than 44 nM/liter. In the most severe cases the pH may fall as low as pH 6.9 or even pH 6.8.

2) Alkalosis[1] The blood and extracellular fluids may on the other hand become more alkaline (or less acid) than normal. Here the pH is elevated, again to a greater or less degree, above pH 7.44 and the hydrogen ion concentration is less than 36 nM/liter. In the most severe cases the pH of the blood may be as high as pH 7.8.

It is evident that the determining factor for the pH of the blood is the ratio of $[HCO_3^-]$ to $[CO_2]$. When this ratio is decreased the pH of the blood is lowered and a state of acidosis is present, when it is increased the pH is elevated and an alkalosis has developed. The ratio may be decreased or increased by changes either in the concentration of the carbonic acid (the denominator) or of the bicarbonate ion (the numerator) component of this buffer fraction. pH changes in the blood caused primarily by alterations in carbonic acid levels are, as a general rule, brought about by alterations in respiratory function and can, therefore, be referred to as being 'gaseous' or 'respiratory.' On the other hand, pH changes which are primarily due to alterations in the bicarbonate ion concentrations are usually associated with metabolic disturbances and therefore can be called either "metabolic" or 'nonrespiratory.'

With these observations in mind the possible primary abnormalities in hydrogen ion regulation may be listed as in Table 4-17.

Although the various activities briefly described above are reasonably successful in preventing large and dangerous changes in H^+ concentrations the compensation is rarely complete. In most cases of acidosis, other than in those of mild degree, the pH of the blood will be displaced to a greater or less degree below the lower range of normal. Similarly, in most cases of alkalosis the pH will be elevated to levels which are more or less above the upper limit of the normal range (Table 4-18).

TYPICAL ABNORMAL STATES

The four major types of acid-base disturbances may be briefly described as follows.

Respiratory Acidosis *Primary Carbon Dioxide Excess, Carbon Dioxide Retention, Hypercapnia.* In emphysema, pulmonary fibrosis, acute pulmonary edema, pneumothorax and in other pulmonary diseases, in conditions such as abdominal distention, gross obesity, bulbar poliomyelitis, opiate poisoning and in situations in which the air being breathed already has an elevated CO_2 content, the ability of the lungs to excrete even normal amounts of carbon dioxide may be grossly impaired. As a result the P_{CO_2} of the body fluids may become markedly elevated. The $[HCO_3^-]/[CO_2]$ ratio and the pH of these fluids will consequently be depressed below the normal values because of increase in the denominator of the buffer pair and a state of *respiratory acidosis* will be produced. The extent to which these changes do occur will be minimized by the following reactions:

[1] Some will disagree with using the terms "acidosis" and "alkalosis" in the sense in which they are employed in this chapter and would prefer, perhaps, to use the terms "acidemia" and "alkalemia" (or "basosis") in their place. "Acidosis" and "alkalosis" have long been used by clinicians and they are consistent with the initial direction of the change taken by the H^+ concentration in response to the primary clinical disturbance. It does not seem valid to define these terms, as some have done, solely on the basis of the concentration of plasma "bicarbonate," "total CO_2," or "total buffer base," because the determination of any one of these by itself gives no indication at all of the $[HCO_3^-]/[H_2CO_3]$ ratio and it is this *ratio* rather than the absolute amounts of either (or of the sum of both) that determines whether the plasma is more acid or more alkaline than normal. For further discussions of these and other words used in the terminology of acid-base disturbances, see Creese et al. (1962) and also the proceedings of the "Conference on Current Concepts of Acid-Base Measurements," New York Academy of Sciences, November, 1964 (Ann. N. Y. Acad. Sci., 133, 1966).

Table 4 17
Illustrative examples of possible primary abnormalities in hydrogen ion regulation (if uncorrected or uncompensated)

Clinical Condition	Biochemical Abnormality	Plasma $\frac{[HCO_3^-]}{[H_2CO_3]}$ Ratios		
		mM/liter	Numerical	pH
Normal	None	26 mM/1 3 mM	20/1	7 4
Acidosis				
I "Gaseous" or "respiratory" acidosis	Primary CO_2 excess $\frac{[HCO_3^-]}{[CO_2]}$ increased	26 mM/2 6 mM	10/1	7 1
II "Metabolic" or "nonrespiratory" acidosis	Primary bicarbonate deficiency $\frac{[HCO_3^-]}{[CO_2]}$ decreased	13 mM/1 3 mM	10/1	7 1
Alkalosis				
I "Gaseous" or "respiratory" alkalosis	Primary CO_2 deficiency $\frac{[HCO_3^-]}{[CO_2]}$ decreased	26 mM/0 65 mM	40/1	7 7
II "Metabolic" or "nonrespiratory" alkalosis	Primary bicarbonate excess $\frac{[HCO_3^-]}{[CO_2]}$ increased	52 mM/1 3 mM	40/1	7 7

a) Whenever CO_2 tends to accumulate, the respiratory center, if not depressed, will stimulate an increased degree of ventilation if this is possible, and the excess CO_2 will be excreted Often, however, as in conditions such as those mentioned above, the respiratory mechanisms are not able, for one reason or another, to respond to a sufficient degree to the stimulus and the P_{CO_2} remains elevated The respiratory center soon loses its sensitivity to increased CO_2 tensions and a chronic state of carbon dioxide retention is established In severe cases the CO_2 accumulation may continue to elevate the CO_2 concentration in the body fluids and a CO_2 narcosis may eventually develop Mechanical aids to respiration may be necessary to control the accumulation and to prevent the onset of this dangerous complication

b) The buffering actions of the various body buffer systems, including those inside the tissue cells, help to limit the extent of the $[H^+]$ increase associated with the accumulation of the excess carbonic acid and at the same time they cause the HCO_3^- concentration in the plasma and other extracellular fluids to be increased to some degree.

c) The kidneys increase the excretion of hydrogen ions and, by almost completely reabsorbing the bicarbonate ions from the glomerular filtrate and making new bicarbonate ions which are added to the circulation, the renal tubular cells increase the bicarbonate ion content of the body fluids so that the lowered $[HCO_3^-]/[CO_2]$ ratio and the depressed pH are raised towards more normal values

In the acute cases of severe respiratory acidosis there may be insufficient time for the renal mechanisms to compensate for the increased CO_2 retention and the pH may, therefore, be quickly reduced to dangerously low levels In the more chronic cases, the plasma bicarbonate concentration may be elevated to levels high enough to return the pH of the blood close to the normal range Although the pH in the blood of these patients may be almost normal or at most, only slightly depressed, the total plasma CO_2 content and plasma bicarbonate will be much increased Overly enthusiastic ventilation by mechanical means could, by removing some of the excess CO_2, readily convert the disturbance in such a patient into one of respiratory alkalosis

Metabolic Acidosis. *Nonrespiratory aci-*

Table 4 18
Illustrative examples of possible results of complete or partial compensation on $[HCO_3^-]/[CO_2]$ ratios (normal 26 mM/1 3 mM or (20/1)) and pH (normal = 7 40) in the four major types of acid-base disturbances

	Ratios									
	If uncompensated (unlikely)			If fully compensated (unlikely)			If partially compensated (usual)			
	mMolar	Numerical	pH	mMolar	Numerical	pH	mMolar	Numerical	pH
A Acidosis										
I Respiratory (Denominator ↑)	26 mM/2 6 mM	10/1	7 1	52 mM/2 6 mM	20/1	7 4	46 mM/2 6 mM	17 7/1	7 35
II Metabolic (Numerator ↓)	13 mM/1 3 mM	10/1	7 1	13 mM/0 65 mM	20/1	7 4	13 mM/0 8 mM	16 2/1	7 31
B Alkalosis										
I Respiratory (Denominator ↓)	26 mM/0 8 mM	32 5/1	7 61	16 mM/0 8 mM	20/1	7 4	22 mM/0 8 mM	27 5/1	7 51
II Metabolic (Numerator ↑)	39 mM/1 3 mM	30/1	7 58	39 mM/1 95 mM	20/1	7 4	39 mM/1 5 mM	26/1	7 52

dosis. Primary bicarbonate deficiency The addition of excess amounts of non-volatile or fixed acids to the body fluids will produce a greater or less degree of metabolic acidosis These acids may have their origin from medication (e g ammonium salts) from overproduction (acetoacetic acid in uncontrolled diabetes mellitus lactic acid in severe hypoxia, formic acid in methanol poisoning) or from inadequate renal excretion of acids produced in normal metabolism as in patients with chronic nephritis or renal tubular acidosis The hydrogen ions of these fixed acids are taken up by bicarbonate ions and other extracellular and intracellular proton acceptors, but by this action the concentration of bicarbonate ion and other available buffer anions may be considerably reduced A similar reduction in the concentration of bicarbonate ion will be produced by the actual loss from the body of this ion by diarrhea or drainage from a biliary, pancreatic or intestinal fistula By any of these means the $[HCO_3^-]/[CO_2]$ ratio is decreased because of a decrease in the numerator of the buffer pair and the pH of the body fluids is correspondingly lowered These changes are compensated in part by an increased pulmonary ventilation, which, by 'blowing off' CO_2, decreases the carbonic acid concentration of the body fluids and thus, by lowering the denominator, helps to elevate the depressed $[HCO_3^-]/[CO_2]$ ratio and the pH of these fluids towards more normal values The kidneys also play their part for not only do the renal tubular cells increase their hydrogen ion excretion in the form of ammonium salts and titratable acid but they also aid in the restoration of the lowered plasma bicarbonate ion by almost completely reabsorbing this ion from the glomerular filtrate, and by forming new bicarbonate ion which is added to the circulating blood

Respiratory Alkalosis. *Primary CO_2 deficiency* Excessive hyperventilation, due, not to an elevated P_{CO_2} or increased hydrogen ion concentration, but to voluntary effort, hysteria, anoxic anoxemia, central nervous system disease, or to toxic stimulation of the respiratory center (as in salicylate poisoning), will cause an excessive excretion of carbon dioxide from the body, a decrease in the CO_2 tension and carbonic acid content of the body fluids and, because of the decrease in the denominator, an elevation of the $[HCO_3^-]/[CO_2]$ ratio to levels

higher than the normal 20 to 1 The pH of the body fluids will be elevated and a respiratory alkalosis established In compensation, the decrease in $[H^+]$ of the body fluids will be partly buffered by protons derived from the buffer systems of the extracellular and cellular compartments The renal tubules conserve H^+ by ceasing to excrete ammonium ions and titratable acid Reabsorption of bicarbonate ion from the glomerular filtrate is sharply diminished and considerable amounts of bicarbonate ion are excreted in the urine The plasma $[HCO_3^-]/[CO_2]$ ratio and pH are thus lowered towards normal by decreasing the numerator and at the expense of some of the extracellular bicarbonate In respiratory alkalosis the plasma bicarbonate concentration may be decreased to below the normal range

Metabolic Alkalosis. *Primary bicarbonate excess, Nonrespiratory alkalosis* This type of alkalosis can result from the ingestion of excessive amounts of sodium bicarbonate (or of salts such as sodium citrate or sodium lactate which are transformed by metabolic oxidation into sodium bicarbonate) for therapeutic reasons, or it may be caused by the loss of H^+ ion from the body by the vomiting of hydrochloric acid containing gastric juice In each of these cases the plasma bicarbonate concentration, the $[HCO_3^-]/[CO_2]$ ratio, and the plasma pH all become elevated and a metabolic alkalosis results The increased amount of cation is buffered by protons from extracellular and intracellular buffers The elevated pH depresses the respiratory center and a compensating retention of CO_2 by the lungs may occur to some degree But the simultaneous decrease in oxygen intake, caused by the decreased ventilation, prevents this type of compensation from being very effective Hydrogen ions are conserved by the kidney and some of the elevated plasma bicarbonate is excreted into the urine In the presence of the depressed H^+ concentration in the tubular fluid a reabsorption of sodium by the distal renal tubules is accomplished by exchange for secreted K^+ instead of H^+, and this may lead to an increased K loss in the urine and to a depletion of body potassium If not replaced from exogenous sources, the loss of potassium from the extracellular fluids is partly replaced by potassium liberated from tissue cells in exchange for extracellular H^+ and Na^+ This transfer of H^+ from the extracellular fluid to the cells accentuates the extracellular alkalosis and tends at the same time to cause an intracellular acidosis Similarly, in other cases of potassium depletion, no matter how produced, the loss of H^+ from the extracellular fluid to the cells in exchange for a part of the cellular K decreases the $[H^+]$ of the extracellular fluids and causes the development of an extracellular alkalosis Also, in the potassium depleted individual, the low potassium content of the distal renal tubular cells causes the Na^+ reabsorption occurring at this site to take place mainly in exchange for secreted H^+ instead of K^+ In such cases the urine may be acid (paradoxical aciduria) in spite of the extracellular alkalosis, and the loss of H^+ produced in this way may tend to aggravate the alkalotic state In these cases too, it may be possible to relieve the alkalosis only by supplying the missing potassium

"Mixed" Acid-Base Disturbances Although the four different major types of acid base disturbances described above have been dealt with as distinct and separate entities, it must be remembered that "mixed" types are not uncommon and that a primary abnormality of one kind may readily be superimposed upon one of a different variety A patient with respiratory acidosis from pulmonary emphysema may, for example, develop a metabolic acidosis from uncontrolled diabetes, or a potassium depletion metabolic alkalosis from large doses of steroids used in the treatment of an attack of status asthmaticus, or a patient with a metabolic disturbance may develop an acid-base upsetting abnormality of respiration

Table 4 19
Biochemical abnormalities of the blood plasma in the four principal types of H^+ disturbances

	pH	P_{CO_2}[a]	HCO_3^-	$\frac{[HCO_3^-]}{[CO_2]}$
		mm Hg	*mM/liter*	*Numerical ratio*
Normal	7 4	43	26	*20/1*
Respiratory acidosis	↓	↑	↑	↓
Respiratory alkalosis	↑	↓	↓	↑
Metabolic acidosis	↓	↓	↓	↓
Metabolic alkalosis	↑	↑	↑	↑

[a] mM/liter $[CO_2] = 0\ 0301 \times P_{CO_2}$ at 38° C

The likelihood that such a 'mixed' picture is present will frequently be obvious or be suggested by the history and physical examination The biochemical upset however may be quite complicated, since depending upon the nature of the disturbances, some of the effects of one may either be partially cancelled out or greatly augmented by those of the other In these "mixed" conditions, it is particularly necessary to know not only the 'total CO_2 content" but also the pH and P_{CO_2} of the blood as well In addition, the concentrations of sodium, potassium, and other constituents of the blood should also be determined if a full and correct assessment of the actual acid-base status of the individual is to be made

Investigation and Assessment of Disturbance in H^+ Regulation

The direction of the deviations from normal shown by the more important biochemical characteristics of the blood in each of the four major types of acid-base disturbance are set forth in Table 4 19 As the data in this table show, a depressed pH would be found in both the respiratory and metabolic type of acidosis a lowered P_{CO_2} could indicate the presence of either a respiratory alkalosis or a metabolic acidosis, and an elevated plasma bicarbonate concentration could be a characteristic of either a respiratory acidosis or a metabolic alkalosis Although, in most cases, the nature of the responsible clinical disturbance would suggest the type of acid-base abnormality present,

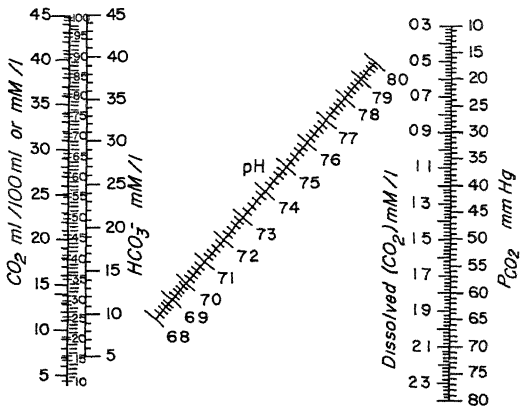


Fig 4 41 Nomogram for graphic determination of serum or plasma values by the Henderson Hasselbalch equation A straight line through given points on any two scales cuts the other two scales at points indicating simultaneously occurring values (From McLean 1938)

```
AC D-BASE EVALUATION   MAY 15, 1969    5:51 PM

SERUM ELECTROLYTES (MEQ/L):

   NA      =
   CO2     =  24
BLOOD PH   =
BLOOD PCO2 =

 IN THE ABSENCE OF A MEASURE OF BLOOD PH OR PCO2 WE CANNOT DETERMINE
WHETHER THE ELEVATED PLASMA BICARBONATE CONCENTRATION IS DUE TO
COMPENSATORY ACIDOSIS, METABOLIC ALKALOSIS, OR TO A MIXED ACID-BASE
DISORDER
DO YOU WISH TO
   1) ASSUME THAT THE DISORDER IS ENTIRELY RESPIRATORY
   2) ASSUME THAT THE DISORDER IS ENTIRELY METABOLIC
   3) SUPPLY THE COMPUTER WITH BLOOD PH OR BLOOD PCO2
PLEASE ENTER A NUMBER> 2

PATIENT'S WEIGHT (IN POUNDS)     = 148

DO YOU HAVE A COMPLETE SET OF SERUM ELECTROLYTES>          NO

IT WOULD PROBABLY BE USEFUL TO OBTAIN THEM

     IF WE ASSUME THAT THE ACID-BASE DISTURBANCE IS ENTIRELY METABOLIC
THE COEFFICIENT OF 36 MEQ/L IMPLIES THAT A PH MUST BE 7 AS UNITS AND
PCO2 46 MM HG RESPECTIVELY THESE ARE THE VALUES THAT WOULD BE COMPATIBLE WITH UNCOMPLICATED METABOLIC ALKALOSIS

     WHILE AWAITING A COMPLETE SET OF ELECTROLYTES IT SHOULD BE
RECALLED THAT METABOLIC ALKALOSIS IS MOST COMMONLY CAUSED BY:
   1) VOMITING OR GASTRIC ASPIRATION
   2) DIURETIC THERAPY (THIAZIDES, MERCURIALS, ETHACRYNIC ACID OR
      FUROSEMIDE)
   3) HYPERADRENALISM (CUSHING'S SYNDROME, PRIMARY OR SECONDARY
      ALDOSTERONISM, ADRENAL STEROID THERAPY)
   4) EXCESSIVE ALKALI INTAKE (ALKALOSIS USUALLY TRANSIENT)
   5) POST HYPERCAPNIA (PARTICULARLY IF CHLORIDE INTAKE IS SMALL)
   6) HYPERCALCEMIA (POSSIBLE - DATA NOT CLEAR)
   7) BARTTER SYNDROME (HYPOKALEMIA, HYPOCHLOREMIA AND METABOLIC
      ALKALOSIS WITH, IN SOME INSTANCES, HYPONATREMIA, SHORT STATURE
      AND MENTAL RETARDATION)

     IF HYPERADRENALISM AND HYPERCALCEMIA ARE ABSENT THEN IT SHOULD
BE POSSIBLE TO CORRECT THE METABOLIC ALKALOSIS BY ADMINISTERING
SUFFICIENT CHLORIDE TO REPLACE PREVIOUS LOSSES AND TO ALLOW A
URINARY CHLORIDE EXCRETION OF AT LEAST 10 TO 20 MEQ PER DAY

THE TOTAL CALCULATED CHLORIDE DEFICIT IS APPROXIMATELY 130 MEQ
IT IS SUGGESTED, HOWEVER, THAT IN ADDITION TO REPLACING KNOWN CHLORIDE
LOSSES, 100 MEQ OF SODIUM, POTASSIUM, OR ARGININE CHLORIDE BE
GIVEN INTRAVENOUSLY DURING THE NEXT 24 HOURS, AND THAT SERUM ELECTROLYTES AND BLOOD PH BE MEASURED AT LEAST DAILY UNTIL SIGNIFICANT
IMPROVEMENT IN THE ACID-BASE ABNORMALITY OCCURS

     THANK YOU FOR REFERRING THIS INTERESTING PATIENT TO US.

REFERENCES:

1) GOLDING, R M., CANNON, P J., HEINEMANN, H O., AND FISHMAN, A P
   RESPIRATORY ADJUSTMENT TO CHRONIC METABOLIC ALKALOSIS IN MAN
   J CLIN INVEST., 47 188, 1968

2) BEESON, P B., AND MCDERMOTT, W.
   CECIL-LOEB TEXTBOOK OF MEDICINE
   12TH EDITION (1967) P 763

3) KASSIRER, J P., BERKMAN, P M., LAWRENZ, D R., AND SCHWARTZ, W B
   THE CRITICAL ROLE OF CHLORIDE IN THE CORRECTION OF HYPOKALEMIC
   ALKALOSIS IN MAN
   AMER J MED 38 172, 1965.

<<>
```

Fig 4 42 Entry of patient's data and computer-generated evaluation Underlined information was typed by the physician everything else by the computer Note that the initial potassium entry was rejected as being unreasonable and that the computer supplied the P_{CO_2} when the physician pressed only the enter button The bicarbonate deficit is calculated on the basis of two-thirds of total body water but the suggested dose (usually less than the deficit) is rounded off to the nearest ampoule when the patient's weight exceeds 50 kg Reproduced from Bleich (1969)

it is obvious that it is not possible to be certain of the exact diagnosis on the basis of the determination of any single biochemical characteristic of the blood plasma, such as the pH or the bicarbonate ion concentration, by itself alone A complete interpretation of the situation in any given patient will require accurate information about each of the three responsible variables: the pH, the P_{CO_2} (H_2CO_3 = 0 0301 × P_{CO_2}), and the bicarbonate ion concentration

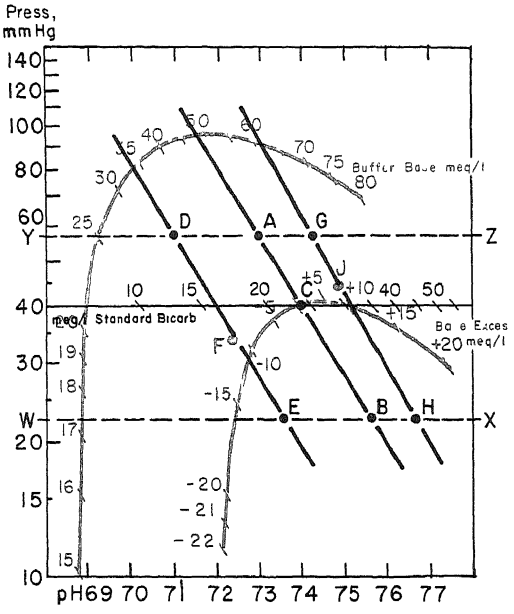


Fig 4 43 The Siggaard-Andersen Nomogram The abscissa is pH and the ordinate is a logarithmic scale of P_{CO_2} Included also are buffer base base excess and standard bicarbonate scales from which these blood values may be determined Buffer lines for three blood samples are shown representing normal acid-base balance (*ACB*) mild metabolic acidosis (*DFE*) and mild metabolic alkalosis (*GJH*) To illustrate the interpolation method of P_{CO_2} determination consider the normal subject (buffer line (*ACB*) The actual pH of the subject's blood is first measured and found to be 7 40 Then small samples of blood are equilibrated with two tensions of carbon dioxide in this example 22 5 and 57 mm Hg corresponding to horizontal broken lines *WX* and *YZ* The pH s of these equilibrated blood samples are measured and found to be 7 566 (for P_{CO_2} 22 5) and 7 302 (for P_{CO_2} 57) these values are plotted on the nomogram (points *A* and *B*) and connected by a straight line which is the buffer line for this blood sample The actual pH (7 40) is then located on this buffer line (point *C*) and the corresponding actual P_{CO_2} value (40) read from the ordinate scale On the acidosis and alkalosis buffer lines shown points *D E G*, and *H* represent equilibrated values and points *F* and *J* represent the actual pH-P_{CO_2} points This illustration is reproduced with the permission of the Radiometer Corporation and Dr O Siggaard-Anderson Copenhagen Figure and legend reproduced from Sharp (1969)

of the arterial plasma Only when the values of all three are known will it be possible to understand accurately the true nature of the disturbance

An inspection of the Henderson-Hasselbalch equation

$$pH = pK' + \log([HCO_3^-]/[CO_2])$$

indicates that if the pK′ of the system is known and if any two of the other three unknowns (pH, concentration of base, concentration of CO_2) can be measured, the third unknown can then be calculated (Fig 4 41) With the development of modern measuring devices it is now possible to measure all three with great accuracy provided the necessary care is taken to collect the blood in such a way that its contents are not disturbed, and to carry out the measurements under the exact conditions of temperature, and O_2 and CO_2 tension which actually exist in the blood

In recent years, the medical use of computers has been increasing An excellent example is the application to the assessment, diagnosis, and treatment of acid-base disturbances This is illustrated in a sample output (Fig 4 42) of a program developed by Bleich (1969) Laboratory results and other data are simply typed, as asked for by the program, on a typewriter console An off-shoot of computer use has been the development of "self-instruction programs" on various subjects in book form The reader is referred to a very fine publication on acid-base physiology by Winters, Engel, and Dell (1967)

Of more widespread use, however, is the excellent Siggaard-Andersen nomogram (Fig 4 43) The nomogram is used with pH measurements of the blood in the Astrup method of determination of P_{CO_2} Whole blood measurements enable graphic estimation of P_{CO_2}, standard bicarbonate and total buffer base (BB) of the blood (Singer and Hastings, 1948), the last being essentially equal to the sum of bicarbonate and proteinate It also enables quick determination of the base excess (BE) of the blood which is the buffer base (BB) found minus the normal buffer base of the blood in mEq /liter Negative values of BE indicate a situation in which a deficit of buffer base exists.

Although the computer is most helpful and nomograms simplify manual calculations, it must be emphasized that a thorough understanding of the interplay of acids and bases in the body's control of hydrogen ion concentration is essential and without substitute

22 FLUID DISTRIBUTION AND EXCHANGE

Measurement of Fluid Volumes

The total volume of water in the body varies among species, in man it has a mean value of approximately 65% of the body weight, or roughly one-third (It is about 65% also in the dog, and about 73% for guinea pigs and rabbits) It is difficult to establish proper mean values for body water, because its total volume is related to the mass of the lean tissues of the body rather than to the body weight In an obese person it is a much lower percentage of the body weight than in one who is lean The values, therefore, vary widely both between persons and different species of animal in accordance with their fatness or leanness

The total volume of the body water can be determined by injecting into the blood stream a known amount of water containing deuterium or tritium, or antipyrine which becomes uniformly distributed throughout the body fluids (intracellular and extracellular), and then determining its concentration in a sample of serum Tritium-labeled water may be more conveniently measured than deuterium-labeled water because of the radioactivity of tritium A further method using *N*-acetyl-4-aminoantipyrene has also been described

The *extracellular fluid*, which comprises the blood plasma, the tissue or interstitial fluid, lymph and the fluid in the serous cavities, amounts to about 24% of the body weight, the plasma water constitutes only about 4% of the body weight in man The fluid within the cells, the *intracellular fluid* amounts to 41% of the weight of the body, or about two times the extracellular fluid The skeletal muscles contain about half, the skin about one-fifth and whole blood only about one-tenth of the total body water (Table 4 20)

Substances, such as sucrose, mannitol, inulin, thiocyanate, and to a large extent chloride, when introduced into the body, become uniformly distributed throughout the extracellular fluids but do not enter the cells in important amounts and, since they are not metabolized and are not excreted too rapidly, can be employed to estimate the volume of the extracellular fluid in the living animal When, for example, a known amount of thiocyanate is injected into the blood stream and sufficient time allowed for equal distribution to occur, the volume of the extracellular water can be calculated from the concentration of the solute in a sample of serum Radioactive sodium or chloride may also be employed for such determinations, but since these solutes are not excluded from all types of cells they are not as dependable as thiocyanate or sucrose, especially if absolute values are sought The extracellular fluid is visualized as occupying a space of the determined volume, and according to the solute used in the estimation, it is customary to speak of the *thiocyanate space*, *sucrose space*, etc It is well to appreciate that some of the methods for determining extracellular fluid volume, because they measure only the volume of fluid into which the test substances can freely diffuse, exclude a certain amount of the fluid that is outside cells, such as eye humors, cerebrospinal fluid, and secretions of glands, whereas these fluids are usually included in the estimation of total body water

The values for the *total body water* and for the *extracellular water* having been determined by the methods just outlined, the *intracellular water* volume can be found by subtraction (see Table 4 21 for percentages) Intracellular water under ordinary conditions is a rather fixed quantity, a loss of one-fourth of the body water is usually fatal

Under physiological conditions the interstitial fluids show the greatest changes In dehydration the proportion of the total water lost from extracellular and intracellular compartments varies with the manner in which the negative water balance is produced When, for

Table 4 20
Water content of various tissues

Tissue	Water Content
	%
Muscle (striated)	75
Skin	70
Connective tissues	60
Adipose tissue	20
Bone (marrow-free)	25–30
Blood	
Plasma	90
Cells	65
Kidney	80
Liver	70
Nervous tissue	
Gray matter	85
White matter	70

Table 4 21
Distribution of body water

Compartment	Percentage of Body Weight
A Extracellular	
In plasma	4
In interstices	15
In hollow viscera, eye, etc	5
B Intracellular	41

example, the dehydration is caused by sucrose diuresis, about 85% of the excreted water is derived from extracellular and 15% from intracellular sources Whereas, in dehydration resulting from water deprivation, the water loss is from 57 to 67% intracellular

Water Balance

In health, except when new tissue is being formed, the body's intake of water obviously must balance the output When the output exceeds the intake, the body's water content is reduced and the body is then said to be in negative water balance, dehydration results When, as during growth, convalescence from an acute illness or in pregnancy, new tissue is being formed, or for a time after a subject has been placed upon a reducing diet, the water balance is positive, the intake exceeding the output, i e , water is retained

The antidiuretic principle of the hypophysis constitutes part of the mechanism regulating the volume of body water In dehydrated states, increased amounts of the pituitary hormone are excreted in the urine (sect 5)

WATER INTAKE

Body water is replenished in two main ways, *a*) by the ingestion of liquids, semisolid, and "solid" food (cooked lean meat, for example, is from 65 to 70% water), and *b*) by the water formed in metabolism through the oxidation of the hydrogen of the food, or of the body tissues themselves Table 4 22 gives the quantities of water produced by the metabolism, respectively, of the three main food stuffs and of alcohol Water is also formed in the tissues through the polymerization or synthesis of various compounds, i e , through a metabolic process the reverse of hydrolysis An ordinary mixed diet yields as a result of oxidative processes from 300 to 350 gm of water daily, or about 14 gm /100 kcal When no food or drink is taken, the body materials themselves are utilized for this purpose, the glycogen, protein, and fat supplying important quantities of water The camel's hump, for instance, which is largely composed of fat, is a reservoir for large amounts of water, and the clothes moth kept in a desiccator and fed upon perfectly dry food lays eggs which are 80% water

For the adult, the amount of water from all sources and under ordinary circumstances which must be ingested daily is around 2500 ml or about 1 ml /kcal of food intake This usually means that about 1000 ml of water as such or in beverages must be drunk in order to maintain the water balance

The water intake under conditions of average temperature, humidity, and diet is summarized in the following Table 4 23 The volume of body water is held constant through a nicely balanced adjustment between intake and output The intake is regulated mainly by the sensation of thirst (sect 9, chap 7)

WATER OUTPUT

Water is lost from the body in the feces, urine and saliva, and by the evaporation of water from the skin and lungs The daily loss through these several channels is given in the following Table 4 24 for a man of average size at light occupation in a temperate climate Under usual conditions of air temperature (23 to 25° C), humidity, and diet, the heat lost from the lungs and the surface of the body by the evaporation of water amounts to about 24% of the total heat production The measurement of this *insensible water loss* under standard conditions may therefore be employed as a

ble 4 22
ater of oxidation

100 gm	of fat	yield 107 1 gm of water
100 gm	of starch	yield 55 1 gm of water
100 gm	of protein	yield 41 3 gm of water
100 gm	of alcohol	yield 117 4 gm of water

ble 4 23
ily water intake

Source	Vol /day
	ml
olid and semisolid food	1200
xidation of food	300
rinks (water, milk, coffee, beer, etc)	1000

able 4 24
aily water loss

Site of Loss	Vol /day
	ml
kin (at average temperature and humidity)	500
Expired air (at average temperature and humidity)	350
Jrine	1500
Feces	150
Total	2500

basis for the determination of the basal metabolic rate

The water lost through the skin and lungs varies greatly with the temperature and relative humidity of the atmosphere and with the extent of the muscular exercise indulged in At ordinary temperatures slight secretion by the sweat glands is not perceived since the sweat evaporates as quickly as it is formed This *insensible perspiration*, as it is called, includes the loss of a greater amount of water by evaporation from the moist tissues beneath the skin, this loss is quite apart from the actual secretion of visible sweat The diffusion of water through the skin and evaporation from the surface under ordinary comfortable conditions of room temperature and humidity is around 1 mg/cm ^{2} of skin surface in a period of 10 min The rate of diffusion is little different for living or dead skin and whether sweat glands are present or absent When the air is hot or the body temperature raised, the rate of evaporation of water from the skin is much more rapid In hot climates the daily secretion may amount to 3000 ml daily, and in very torrid atmospheres it may be as much as 10 liters When heavy work is done in a hot environment, sweat may be secreted at the rate of 2 liters/hr This necessitates the drinking of an equal quantity of fluid in order to maintain the normal water content of the body since the intake must equal the output

At ordinary temperatures the inspired air contains negligible quantities of water whereas the expired air is almost saturated with moisture Any condition which increases the pulmonary ventilation therefore increases the water lost by this route

REGULATION OF VOLUMES AND OSMOLARITY

The isotonicity of the body fluid which depends mainly upon its concentration in sodium and chloride is maintained constant by the retention or elimination of both water and electrolytes the kidneys playing the primary role in this regulation Just how this regulation is achieved is not clearly understood The hypothalamus is important in controlling both the liberation of antidiuretic hormone by the posterior pituitary gland, and water intake (sect 9, chap 7) There must be, however, some means of regulating the absolute amount of body water The basic factor in this control may be the regulation of sodium excretion by the kidney (sect 5, chapt 1) Circulatory changes are also important Factors leading to a reduction in circulating blood volume, venous return, or cardiac output appear to reduce the excretion of salt and water and increase the water intake, whereas factors increasing blood volume or the venous return and cardiac output increase the excretion of salt and water and inhibit thirst

Changes in the diet can bring about changes in the total body water Protein is laid down in the body with water (about 3 gm of water per gram of solid) During growth or convalescence from wasting diseases, retention of water therefore occurs, i e , the intake of water, including that derived from solid food, exceeds the output Fat is laid down with a minimal amount of water (only that in the protein of the connective tissue framework), the deposition of glycogen is accompanied by a small storage of water Water retention therefore follows a sudden change from a high fat to a high protein diet, and to a lesser degree from a fat diet to one high in carbohydrate A change

from a diet high in carbohydrate or protein to one high in fat is followed by the loss of water

Reduction of the caloric value of the diet below the energy requirement is accompanied for a time by the retention of water During the 1st week or so on a reducing diet the subject's weight may for this reason show little or no change, the fat catabolized being replaced by water In a prolonged fast after the fat stores have been depleted, protein is also partly replaced by water, the muscles of animals dying of starvation showing a marked reduction in the proportion of protein

DEHYDRATION

Major Causes of Dehydration *1) Water Depletion or Primary Dehydration a) Simple deprivation of water* from whatever cause shipwreck, desert travel, dysphagia, extreme weakness, mental patients refusing to drink, etc Under such circumstances, though there is an effort to conserve the stores of body water, through a reduction in the amount excreted by the kidney, in the sweat and by other routes, some water is always lost though none be drunk Dehydration occurs more quickly in fever or if the environmental temperature is high

b) Excessive water loss may result from persistent vomiting (e g , pyloric or intestinal obstruction), prolonged diarrhea, or the excretion of large quantities of urine or sweat, especially when accompanied by a restricted water and salt intake In the acute diarrheas of infants, dehydration and loss of weight may occur very rapidly Such water loss is accompanied by serious depletion of electrolytes

In water depletion the osmotic concentration of the extracellular fluid rises, water is drawn from the cells, and both extracellular and intracellular compartments shrink Extreme thirst is experienced

2) Reduction in the Total Quantity of Electrolytes, Salt Depletion, or Secondary Dehydration The electrolytic concentration of the body fluids, both extracellular and intracellular, is maintained constant through the elimination or retention of water. That is, a reduction or increase in the total electrolytes, which comprise chiefly the cations Na^+ (extracellular) and K^+ (intracellular) and the anions, HCO_3^- and Cl^-, is accompanied by a corresponding decrease or increase in the volume of body water. The sum of the cations and anions of course must balance Loss of Cl^- can be made good by the retention of CO_2 and a rise in plasma bicarbonate Excreted cation, however, can be replaced only by basic substances supplied in the food The total concentration of electrolytes in the body fluids is therefore dependent upon the stores of total cation For example, in pyloric or high intestinal obstruction, fluid is secreted in large quantities into the gastrointestinal tract The fluid may be vomited or may collect and remain in the dilated part of the canal above the obstruction In either case the secretion of large quantities of gastric juice entails a loss of blood chloride, and an appreciable quantity of K^+ The chloride depletion causes at first no ill effects, the normal concentrations in electrolytes of the blood and tissue fluids being maintained for a time by the retention of CO_2, and, as a consequence of this, an increase in bicarbonate The compensation for the Cl^- loss may lead, however, to alkalosis which is then countered by an increased excretion of Na^+ in the urine This of course is accompanied by diuresis, marked dehydration results

On the other hand, the continued loss of pancreatic juice to the exterior causes an immediate depletion of cation, plasma bicarbonate is reduced In the adjustment of the acid-base balance the excess of anions is excreted in the urine, this again entails a loss of water Similarly, the ingestion of acid-producing salts causes a depletion of cations, which are used for the neutralization and excretion of the acid radicles Such salts therefore act as diuretics and dehydrating agents

In salt depletion the extracellular fluid is hypotonic, water enters the cells, so that the volume of intracellular fluid is maintained or even increased, whereas the extracellular fluid (especially the interstitial) is reduced

Clinically, the leaching of salt from the body of a seriously ill patient by glucose infusions used to be a not uncommon cause of salt depletion

3) The Injection of Hypertonic Solutions into the Blood Stream When a strong sugar or salt solution is injected, the temporary rise in the osmotic pressure of the blood causes a flow of fluid from the tissues into the vascular system until equilibrium is re-established The blood volume is increased, but is soon returned to normal by the loss of the excess material

into the tissues and its eventual excretion via the kidney and bowels A net loss of body water results

Effects of Dehydration *a) Loss of weight* due to the reduction in tissue water as well as to the actual breakdown of body substance which occurs in the effort to furnish water for the maintenance of physiological processes Fat and carbohydrate stores are first drawn upon for this purpose and later, protein *b) Disturbances in acid-base balance* usually toward the acid side The diminished quantity of circulating fluid (loss of plasma water, anhydremia) and the consequent depression of oxidative processes in the tissues are held responsible for the excessive production of acid metabolites, e g , lactic acid The slowing of the renal circulation also leads to a reduced excretion of urine and the retention of anions, e g , phosphate and sulfate, which under normal circumstances are eliminated *c) Rise in the nonprotein nitrogen of the blood d) Rise in plasma protein concentration* and of chloride though there is no absolute increase There is an absolute increase in blood sugar, especially when the stage of exhaustion approaches *e) Rise in body temperature* as a result of the reduction in circulating fluid (sect 9, chap 7) *f) Increased pulse rate and reduced cardiac output g) Thirst* This occurs in water depletion but not in salt depletion Any fall in the water content of the tissues is reflected in the glandular activities especially of the salivary glands Secretion is suppressed, the mouth and throat become dry In dehydration thirst is extreme and the mouth parched *h) Dryness, wrinkling, and looseness of skin* and a pinched expression to the features result from the loss of subcutaneous fat and of water from the deeper layers of the skin Other manifestations are, reduced intraocular tension and recession of the eyeball and, in infants, depression of the fontanelle *i) Exhaustion and collapse*

WATER INTOXICATION

When an animal is given large quantities of water by stomach tube, especially if urinary secretion is reduced by the administration of pitressin, the tissues become "water-logged," serious symptoms ensue, e g , depression of temperature, vomiting, convulsions, and coma, which shortly end in death Similar effects also follow in man if large quantities of water are given to a patient with nephritic edema or, if in a subject of diabetes insipidus, pituitrin be administered while the water intake is maintained at the usual level The manifestations of water intoxication are believed to be due to the dilution of electrolytes in the body fluids, and the damage caused thereby to the tissue cells

Adrenalectomy reduces the renal response to water drinking, and thus increases the susceptibility to water intoxication This susceptibility is reduced by the administration of desoxycorticosterone or of 17-hydroxy-11-dehydrocorticosterone or of thyroid hormone A similar protective action is exerted by these principles upon normal animals

In water intoxication the protein and chloride of the plasma are diminished, although much of the water retained in the body enters the cells of blood and tissues which become swollen The artificial administration of inordinate amounts of water in the form of glucose solution, especially after surgical operations when there is some tendency toward antidiuresis, may cause serious disturbances in water metabolism The effects of an excess of body water induced in this way can be corrected readily by the administration of hypertonic saline

Blood Volume

Blood volume regulation is largely a question of balance between the fluid within the vessels and in the tissues When conditions arise which tend to lower or raise the volume of blood, counter forces come into play which restore the normal level When circulating fluid is lost, the vessels replenish themselves from the extravascular spaces On the other hand, any tendency for the blood volume to rise is met by a discharge of the excess fluid into the tissues and later from the body in the urine So, a balance is struck, and in health the blood volume is maintained remarkably constant For example, after the intravenous injection of a large quantity of saline, the volume of circulating fluid (although raised temporarily) is brought back toward normal within 30 min or less On the other hand, the loss of whole blood, as by hemorrhage, immediately calls into action processes which tend to replenish the blood volume When an animal is bled to death, at a not too rapid rate, the blood which is withdrawn is found after a few minutes of bleeding to have become diluted—clearly demonstrating the promptness

with which fluid (water and salts) has been absorbed into the vessels

Blood volume estimations are of value in many clinical conditions associated with a loss or gain of fluid by the body, or for the purpose of checking the results of corpuscular and hemoglobin estimations It is also sometimes of interest to know, in experimental investigations and in metabolic studies in man, the *total* amounts of certain blood constituents, e g , protein, calcium, sodium, etc , as well as their concentrations

DIRECT METHOD FOR THE ESTIMATION OF BLOOD VOLUME

The first attempts to measure the total quantity of blood in the body were made upon animals by Welcker (1854) His method consisted in taking a small measured quantity of the animal's blood and diluting it to 1 in 100 with normal saline The animal was then bled, and after the blood had ceased to flow, its vessels were washed out and the muscles minced and extracted with water Water was then added to the collected fluid—blood and washings—until its color matched precisely the tint of the original diluted blood specimen The total collected fluid divided by 100 gave the blood volume

This method was also employed upon decapitated criminals (Bischoff, 1857) in order to obtain a value for the blood volume of man, but since its applicability was obviously very limited, indirect methods were later devised

INDIRECT METHODS FOR ESTIMATION OF PLASMA VOLUME

Several methods are available for the determination of the blood volume during life

Method Using T-1824. The dye method was originally devised by Keith, Rowntree, and Geraghty (1915) The degree of dilution in the plasma of a known amount of dye injected into the circulation is employed as the basis of calculation The color of the stained plasma is compared in a colorimeter with that of a standard dye solution of known concentration There are several qualifications required of a dye before it can be considered suitable for blood volume measurements In the first place, of course, it must be innocuous It must not diffuse too rapidly from the blood stream, it must color only the plasma, and not be adsorbed by the cells of the blood nor by the walls of the blood vessels It must not change color itself after entering the blood, nor cause the liberation of pigment from the red cells (hemolysis), and it must be capable of mixing evenly and thoroughly with the plasma If it lacked these qualifications, the colorimetric readings obviously would be undependable

The dye most commonly employed is a blue dye, T-1824 (Evans blue) Evans blue has the advantage that the error due to any discoloration of the plasma by hemoglobin (hemolysis) is minimized This dye is also eliminated very slowly (4 8 %/hr) from the circulation It remains in the circulation because it combines with protein, although this combination apparently is not strong, for the two can be separated by paper electrophoresis

In making an estimate of blood volume a sample of plasma is obtained before the dye is injected Then a known amount of the dye is injected intravenously and blood samples are obtained at 10-, or 20-, 30-, 40-, and 50-min intervals after the injection The concentration of dye in the plasma is measured Because the dye leaves the circulation to some extent, it is necessary to determine the fall in concentration of the dye with time By extrapolating back to zero time from the curve of concentrations, the concentration of the dye at zero time can be computed and this is used to calculate the blood volume Because T-1824 leaves the circulation slowly, in normal persons it is usually possible to obtain a reasonable estimate of the blood volume from a sample taken 10 min after the dye has been injected, but under conditions when mixing is slow, the extrapolated values are preferable From the concentration of dye in the sample and the amount of dye injected, the dye dilution, and hence the plasma volume can be computed

Method Using ^{131}I. Plasma volume may also be estimated by the injection of albumin "tagged" with ^{131}I This is not affected by hemolysis or lipemia, nor does it stain the tissues, but it also has the disadvantage of requiring information concerning its disappearance rate Because only minute amounts of radioactivity are used in each test, repeat determinations can be made accurately and in rapid succession This technique has been adapted for use at the bedside with the advent of improved equipment, and has largely supplanted the use of Evans blue Measurements of plasma volumes by the use of T-1824 and by means of ^{131}I albumin give similar results

Dextran Method. Another procedure has been developed using a dextran fraction which leaves the circulation very slowly. This method has the advantage of requiring only two blood samples, one of which is a control. It is unaffected by lipemia or hemolysis, employs a colorless material and requires no highly specialized equipment for its estimation.

CALCULATION OF BLOOD VOLUME

These various procedures give an estimate only of the *plasma* volume. In order to obtain the value for the total blood volume, the proportion of cells to plasma must be determined by means of the hematocrit. If the *packed cell volume* thus obtained is, for example, 45%, the plasma volume is therefore 55%. Thus, the total blood volume can be calculated from the plasma volume and the hematocrit reading.

In calculating the blood volume from the plasma volume and the hematocrit reading it is necessarily assumed that the proportion of red cells to plasma is the same throughout all parts of the vascular system, but experimental evidence is strongly against such an assumption. In capillaries the proportion of red cells to plasma may be considerably less or more than in the larger vessels. To eliminate the error of unequal distribution of red cells throughout the circulation, methods have been devised for the estimation of the total circulating red cell volume. These use the dilution of injected red blood cells which have been "tagged" with radioactive iron (^{59}Fe), radioactive chromium (^{51}Cr) or radioactive phosphorus (^{32}P). This method is based on two assumptions—*i*) that none of the radioactive atoms escape again after being taken up by the erythrocytes, and *ii*) that all the labeled cells become mixed with the recipient's cells equally throughout the circulation. Only the principle of the method can be given here, for details the reader is referred to the original papers (Hahn et al., 1942, Gibson et al., 1946). Radioactive chromium is the isotope used most frequently for this purpose.

In order to determine the total blood volume, the plasma volume is measured, for example, by the ^{131}I method, and that result is added to the value obtained for the circulating red cell volume.

NORMAL VALUES FOR PLASMA AND WHOLE BLOOD VOLUMES

In normal subjects, blood makes up roughly 7% of the body weight, with plasma volume a little over 4% and red cell volume about 3%. The plasma volume of adult males is about 40 ml., the red cell volume about 29 ml., and the whole blood volume in the neighborhood of 70 ml./kg. of body weight. The blood volume of a man of average weight (70 kg.) is therefore around 5 liters. The blood volume in terms of body weight shows considerable variation from person to person. Changes in body density alter this relationship of blood volume to body weight. The relationship between blood volume and body height shows a similar variability. The most constant relationship is found between a combination of body weight and the cube of body height.

ALTERATIONS IN BLOOD VOLUME

Major Causes of Reduction of the Blood Volume. *1*) A loss of *whole blood* as in hemorrhage. *2*) *Reduction in the total number of red cells* as a result of increased destruction or diminished production (see anemia, chap. 3). *3*) Loss of *plasma* alone from the vessels as in burns or shock from other causes. *4*) *Loss of blood water*. This is called *anhydremia* and is simply a part of a general dehydration and so results from the same causes as the latter. *Acute exposure to cold* causes a moderate loss of water from the blood to the tissues (chiefly skin, muscle and probably liver), the total water content of the body remaining unaltered. It is not altogether clear by what means this movement of water is brought about. The blood volume of the human subject after 30 min. or so in the *erect posture* is some 15% less than that in recumbency. A fluid of low protein concentration leaks from the vessels of the lower limbs into the extracapillary tissues as a result apparently of the increased hydrostatic pressure in the capillaries of these parts. For this reason, a subject must be in the recumbent position in order to obtain an accurate measurement of blood volume.

Increase in Blood Volume. *1*) *High Temperatures*. In hydrated subjects, exposure to a high temperature leads to an increase in blood volume. This increase reaches a maximum in about 7 days after which it may return toward normal. The extracellular fluid volume increases also, although initially the greater part of this increase is reflected in the elevated plasma volume. In the first few hours after exposure to heat the changes are slight and inconsistent.

2) *Muscular Exercise*. The effect of a bout of

vigorous exercise in man is to decrease the plasma volume The magnitude of this decrease may be exaggerated when measurements are made by the dye method because of a change in the optical properties of the plasma However, the protein concentration of the plasma, and the red cell count increase The increase does not seem to be progressive Some elevation in blood volume may occur with muscular training This can be demonstrated in dogs, but in men accustomed to hard physical work, the red cell volume in relation to height and weight is only slightly increased over that of sedentary individuals

3) *Emotional Excitement* in animals and in man causes an increase in blood volume

4) *Pregnancy* In pregnant women, the blood volume though somewhat variable increases on the average by 20 to 30%, though in the last few weeks of pregnancy tends to fall again There is an increase in plasma volume and in interstitial fluid Associated with this is a retention of sodium Restriction or the administration of salt reduces or increases, respectively, the volume of extracellular fluid The increase in plasma volume in pregnancy, combined with a lower protein concentration of the plasma is responsible, in part at least, for the edema of the lower limbs which commonly occurs in the pregnant state Hemoglobin determinations in the last trimester of pregnancy give a falsely low indication of total red cell mass

5) *Congestive Heart Failure* As a result of the retention of sodium, the extracellular fluid undergoes a pronounced increase

6) *The Administration of Desoxycorticosterone* causes a retention of salt and, as a consequence, a rise in the volume of plasma and interstitial fluid This depends on the salt load Aldosterone is much more active than desoxycorticosterone in its effect on sodium retention It helps to keep in equilibrium the fluid volume and sodium retained, fluid retention varying directly with that of sodium

PATHOLOGICAL STATES ASSOCIATED WITH ALTERATIONS IN BLOOD VOLUME

Reduction. *a) Hemorrhage* (loss of whole blood), *b) Burns* (loss of plasma), *c) Dehydration* (loss of water), *d) Pernicious Anemia* (reduction in red cells, with a moderate increase in plasma), *e) Certain Chronic Anemias Other Than Those of the Pernicious Type* In these the total blood volume, as a rule, is not greatly below normal as calculated upon the basis of weight or of surface area *f) Obesity* The blood volume per kilogram of body weight is much reduced but is normal when considered in relation to the body surface *g) Myxedema* (reduction of both red cells and plasma but mainly of the former)

Increase. *a) Polycythemia Vera* (increase mainly of red cells but also of plasma), *b) Cirrhosis* of the liver (increase of plasma), *c) Leukemia* (increase in white cells and plasma), *d) Splenomegaly* with anemia—Banti's disease (increase in plasma), *e) Hyperthyroidism* (equal increases both in red cells and plasma), *f)* congestive heart failure (increase in plasma)

It should be pointed out that the proportions of red cells and plasma may vary from the normal though the total blood volume remains unaltered With regard to the blood volume and the proportion of cells to plasma there are therefore nine possible blood states

THE EFFECTS OF HEMORRHAGE

When more than 30% of the blood volume is lost rapidly, the body is usually unable to repair the loss unaided and, unless transfusion is resorted to, death results In a healthy man the loss of 30% or less of his blood calls readjusting mechanisms into play which may bring the blood volume back to the normal level within a remarkably short time, 500 ml or so of blood drawn for transfusion purposes are said to be replaced within an hour or so The restoration of the blood to its previous concentration in erythrocytes, however, takes about 7 weeks on the average This time may be shortened considerably by the administration of iron and a diet containing a liberal quantity of high quality protein Fowler and Barer found in a study of 200 blood donors that, after the removal of 550 ml of blood, the average fall in hemoglobin was 2 3 gm /100 ml Regeneration of hemoglobin occurred at the rate of 0 049 gm /100 ml for men and 0 040 gm /100 ml for women per day

The protective mechanisms which automatically come into action after hemorrhage are several and may be divided into two groups—*immediate* or *early* and *delayed*

Immediate or Early Effects. If the loss of blood is large, especially if it is of sudden occurrence and from an artery, there is a prompt fall in blood pressure as a result simply of the

reduction in circulating fluid If not too great the fall in pressure is salutary since it helps to prevent further bleeding A moderate loss of blood, 10% of the total amount, produces little or no drop in pressure This is especially true if the blood is lost gradually and if it comes from a vein i e beyond the peripheral resistance Under these circumstances compensatory mechanisms easily maintain the pressure of blood at its normal height Three main compensations relating to the vascular system occur *a)* reduction in the capacity of the vascular bed, *b)* increased peripheral resistance, and c) fluid entry from the tissues

Reduction in Capacity of the Vascular Bed and Redistribution of the Blood When the flow of blood has been stopped or considerably lessened by a complete or partial closure of the wound in the vessel, the blood pressure, if this had fallen, rises again This is the result mainly of a readjustment of the capacity of the vascular system, whereby it is made to conform more nearly to the lessened volume of blood It is this reduction in the vascular capacity which prevents the initial fall in pressure when the loss of blood is gradual It is effected by the reflex narrowing of innumerable small vessels (vasoconstriction) in regions such as the skin, mucous membranes, intestine, and other parts not immediately essential to life If the hemorrhage has not been too severe the reduction in the capacity of the vascular bed may result in maintenance of an adequate venous return and consequent maintenance of cardiac output and of arterial blood pressure If the hemorrhage is somewhat greater, then the venous return and cardiac output may fall, but the blood pressure may be maintained by an increase in peripheral resistance along with a reduction in the capacity of the arterial vessels If the hemorrhage is still greater, then the arterial blood pressure will fall and may be restored by restoration of blood volume through the entrance of fluid into the blood vessels from the tissues The vascular response is called into play by the underfilled state of the arteries and large veins feeding the heart (see vascular reflexes, sect. 3, chap 8)

These measures, whereby the blood remaining in the vascular system is confined to a smaller space are of the utmost importance, they enable the essential centers in the medulla to be supplied with blood under adequate pressure to sustain their vitality Also a greater quantity of blood than would otherwise be possible is brought to the heart to supply its muscle, fill its cavities and maintain the circulation The withdrawal of blood from the less important parts of the body is responsible, however for some of the characteristic manifestations of hemorrhage notably the pallor of the skin and mucous membranes and the coldness of the body surface The cerebral anemia causes sensations of giddiness or faintness, flashes of light or ringing in the ears (tinnitus)

Delayed Effects *Replacement of the Lost Fluid* This, it has already been mentioned, commences almost upon the instant that the blood is lost, but takes a variable length of time, depending upon the extent of the blood loss to become complete It occurs as a result of the low hydrostatic pressure within the capillaries Fluid from the tissues moves into the vessels and dilutes the blood The corpuscular concentration is therefore reduced and since fluid is entering the capillaries directly from the tissues and not by way of lymphatics, less protein is returned with the fluid from the tissues, and for a short time after hemorrhage the protein content of the plasma is markedly depressed Very soon, however, the concentration of protein in the plasma shows a rise again as a result of the mobilization of protein stores probably from the tissue stores The extreme thirst (sect 9, chap 7) which the subject of acute hemorrhage suffers is the call of the tissues for fluid and indicates that their own fluid stores are being drawn into the underfilled vessels The administration of water will therefore aid the body in recovering its water balance and replenishing the blood volume

Replacement of the Blood Cells finally occurs through the increased activity of the blood-forming organs This takes several days or weeks, the rapidity of the process depending to a large extent upon the nutrition and recuperative power of the individual and upon the diet as well as the amount of blood lost While the repair process is in progress reticulocytes are found in increased numbers in the blood

Interchange of Fluid Between the Tissues and Capillaries

The physical factors which determine the flow of fluid from the tissues into the blood stream as well as in the reverse direction—from the vessels to tissue spaces—are the *os-*

motic and *hydrostatic pressures* of the fluids in the two situations *Osmosis* is diffusion through a membrane separating water—or an aqueous solution—to which the membrane is permeable, from another aqueous solution to whose dissolved materials the membrane is impermeable Such a membrane that permits the passage of water or of some substances in solution, but acts as a barrier to the molecules of certain other substances, is called *semipermeable* Osmosis is explained in section 1, chapter 2

Isotonic, *hypertonic*, and *hypotonic* are terms employed to denote the osmotic activity of a solution as compared to that of some other solution If, for example, two salt solutions of equal strength are separated by a semipermeable membrane neither will develop an osmotic pressure, they are said to be *isotonic* If the solution on one side of the membrane is stronger, the net amount of water passing to this side will therefore be greater and the solution here is said to be *hypertonic* The weaker solution from which there is a net transference of water is termed *hypotonic*

In physiology these terms are used most commonly to indicate the osmotic activity of a solution as compared with that of the body fluids For example, excised organs or sections of tissue upon which experiments are to be performed are kept moist and viable by a saline solution isotonic with the tissue fluid Shrinkage of the cells through the loss of water (as would result from the use of a hypertonic solution) or swelling by endosmosis (as would occur if the solution were hypotonic) is thus avoided Similarly, in order to prevent either shrinkage or swelling of the red blood cells, they must be suspended in a solution, such as plasma, that is isotonic with the fluid in their interiors This subject is reviewed in chapter 6 of this section

The other important factor in the interchange of fluid between the blood and the interstitial (and hence, the intracellular) fluid is the *hydrostatic pressure* within the capillaries, i e , the intracapillary blood pressure and its relation to the pressure in the tissue around these small vessels The concept that fluid exchange across capillary membranes is governed by the net difference between the osmotic and the hydrostatic pressures is known as the Starling (1896) hypothesis or principle.

THE RELATION OF OSMOTIC TO HYDROSTATIC PRESSURE ACROSS THE CAPILLARY MEMBRANE

The walls of the capillaries (generally throughout the body, as already mentioned) permit the free passage of water and substances of small molecular size dissolved in the plasma, e g , glucose, inorganic salts, urea, etc But, the capillary wall is relatively impermeable to the larger molecules such as the plasma proteins, the presence of which is responsible for the difference in osmotic pressure on the two sides of the capillary membrane It is important to recognize that some protein molecules do penetrate capillary membranes Since the albumin has the smallest molecule of the three plasma proteins, it escapes through the vessel in relatively greater amounts than do the globulin and fibrinogen fractions By the use of ^{131}I labeled albumin, it has been found that the amount of plasma protein passing through the capillary wall into the interstitial fluid is considerable, amounting to more than the total protein of plasma each day The rate of return of protein from interstitial fluid is very important in the maintenance of plasma protein levels The lymphatic system plays a major role in this return (chap 9)

The manner in which osmotic and hydrostatic pressures act in regulating the interchange of fluids between the tissues and the muscles may now be seen The blood at the arterial end of a capillary has a pressure, let us say, of 30 mm Hg (this is an arbitrary figure and will vary according to the activity of the part) This is a force driving the water and the dissolved crystalloids through the capillary membrane But the hydrostatic pressure of the tissue fluid on the outer side of the membrane offsets, in part, that within The pressure of fluid in the tissue spaces is difficult to determine but it is considerably less than that in the capillaries It probably varies considerably in different regions, being low in those containing much loose areolar tissue But for illustration let it be assumed to be 10 mm Hg The hydrostatic pressure, therefore, which is effective in forcing fluid out of the vessel (filtration pressure) is only the difference between the pressure within and that on the outside of the vessel, namely, 20 mm Hg But the osmotic pressures of the plasma and tissue fluids

must be taken into account In the plasma it amounts to about 25 mm Hg The tissue fluids have a lower protein content and consequently a lower osmotic pressure The latter amounts to about 15 mm Hg Owing to this disparity of 10 mm Hg between the osmotic pressures of plasma and tissue fluid, there would be a net tranference of water and diffusible substances inwards across the capillary membrane if such transference were unopposed by any other force But a net hydrostatic pressure of 20 mm Hg is set against it at the arterial end of the capillary leaving a balance amounting to 10 mm Hg which serves to filter water, salts, etc, and a little protein out of the plasma Now as the blood flows rather slowly through the capillary and loses water in this way by filtration, the protein concentration and consequently its osmotic activity, rises Coincidently with the progressive rise in osmotic activity the hydrostatic pressure declines, until at the venous end of the capillary there may be a balance of net osmotic pressure over net hydrostatic pressure Water accompanied by freely diffusible materials is therefore absorbed into the circulation This picture of filtration at the arterial end of the capillary and absorption at the venous end is going on continually at the periphery of the vascular bed But it is subject to considerable variation, not only in the same capillary from time to time but in different parts of the peripheral circulation at any moment As explained in chapter 9, when a tissue is active most of the water that escapes from capillaries is returned to the circulation via the lymphatics

Table 4 25
Pressures that move fluid

Blood		Tissue Fluid
Hydrostatic pressure 30 mm Hg		*Hydrostatic pressure* 10 mm Hg
Effective hydrostatic pressure 20 mm Hg	Capillary wall	
Osmotic pressure 25 mm Hg		*Osmotic pressure* 15 mm Hg
Effective osmotic pressure 10 mm Hg		

[a] Driving force → 10 mm Hg (20 – 10)

As the blood flows through the capillary as a result of the passage of water outwards there is a consequent rise in the concentration of protein Moreover the metabolic processes in the tissues bring about changes, whereby larger molecules are being broken down into smaller ones other molecules are removed or built up into larger ones In this way alterations in molecular concentrations and in the diffusibility of the constituents of the tissue fluids with consequent variations in osmotic relationships are occurring ceaselessly The actual exchange of water across the capillary wall is rapid Studies with deuterium and tritium-labeled water indicate that in 1 min an amount equal to the total plasma water is exchanged

Under any circumstance in which the blood volume is increased or diminished, either the hydrostatic pressure or the osmotic pressure or both are altered, and it is through such changes that the blood fluid tends to be restored automatically to its previous level After hemorrhage, for example, the hydrostatic pressure is lowered in the capillary area but the osmotic pressure is unchanged Fluid will therefore be absorbed from the tissue spaces Moreover constriction of the capillaries tends to reduce, dilation to increase the capillary pressure Therefore, the capillary constriction which follows a severe hemorrhage will slow the filtration rate at the arterial end of the capillary and increase absorption of interstitial fluid at the venous end Again, when water is drawn from the blood, in consequence of excessive loss of fluid by the kidney, sweat glands, or bowels, concentration of the plasma proteins will result The increased osmotic pressure of the plasma will then hasten the rate of absorption from the tissues The intravenous injection of large quantities of isotonic saline, on the other hand, will have the twofold effect of diluting the colloids and temporarily increasing the hydrostatic pressure The excess fluid in consequence is rapidly eliminated from the blood stream into the tissues and later through the kidney and bowels

BIBLIOGRAPHY

ALPEN E L AND D CRANMORE Cellular kinetics and iron utilization in bone marrow as observed by Fe^{59} radioautography *Ann N Y Acad Sci* 77 753-765 1959

AMOS D B (ed) Human histocompatibility antigens In *Immunology Society Symposium Fed Proc* 29 2010-2047, 1970

ANKER, H S The biosynthesis of plasma proteins In *The Plasma Proteins*, edited by F W Putnam New York Academic Press 1960 vol II, p 267-307

ANTONINI E Interrelationship between structure and function of hemoglobin and myoglobin *Physiol Rev* 45 123-170, 1965

ANTONINI, E AND M BRUNORI Hemoglobin *Ann Rev Biochem* 39 977-1042, 1970

ARMSTRONG D J B JEPSON C A KEELE AND J W STEWART Pain producing substances in human inflammatory exudates and plasma *J Physiol (London)* 135 350-370 1957

ASHBY W The determination of the length of life of transfused blood corpuscles in man *J Exp Med* 29 267-281 1919

ASHFORD T P AND D G FREIMAN Role of endothelium in initial phases of thrombosis An electron microscopic study *Amer J Path* 50 257-273 1967

ASTER R H Pooling of platelets in the spleen role in the pathogenesis of hypersplenic thrombocytopenia *J Clin Invest* 45 645-657 1966

ATHENS, J W, O P HAAB S O RAAB A M MAUER, H ASHENBRUCKER, G E CARTWRIGHT, AND M M WINTROBE Leukokinetic Studies IV The total blood circulating in marginal granulocyte pools and the granulocyte turnover rate in normal subjects *J Clin Invest* 40 989-995, 1961

BAEHNER R L D J NATHAN AND M L KARNOVSKY Correction of metabolic deficiencies in the leukocytes of patients with chronic granulomatous disease *J Clin Invest* 49 865-870 1970

BECKER, A J E A McCULLOCH AND J E TILL Cytological demonstration of the clonal nature of spleen colonies derived from transplanted mouse marrow cells *Nature (London)* 197 452-454 1963

BENESCH, R AND R E BENESCH Intracellular organic phosphates as regulators of oxygen release by haemoglobin *Nature (London)* 221 618-622 1969

BERK P D J R BLOOMER, R B HOWE AND N I BERLIN The life span of the red cell as determined with labeled bilirubin In Formation and Destruction of Blood Cell edited by T J Greenwalt and G A Jameson J B Philadelphia Lippincott 1970 p 91-108

BERLIN, N I, T A WALDMANN, AND S M WEISSMAN Life span of red blood cell *Physiol Rev* 39 577-616 1959

BERT P (1878) *La Pression Barometrique Recherches de Physiologie Experimental* translated by M A Hitchcock and F A Hitchcock Columbus Ohio College Book Co 1943

BEST C H C COWAN, AND D L MacLEAN Heparin and formation of white thrombi *J Physiol (London)* 92 20-31 1938

BIGGS R AND A S DOUGLAS Thromboplastin generation test *J Clin Path* 6 23-29, 1953

BIGGS R AND R G MACFARLANE *Human Blood Coagulation and Its Disorders* Ed 3 Oxford Eng Blackwell, 1962

BILLINGHAM, R E L BRENT T B MEDAWAR Quantitative studies of tissue transplantation immunity II The origin strength and duration of actively and adoptively acquired immunity *Proc Roy Soc (London) [Biol] Ser* B 143 58-80 1954

BISCHOFF T L W Abermalige Bestimmung der Blutmenge bei einem Hingerichteten *Z Wiss Zool* 9 65-72 1857

BISHOP, C R, J W ATHENS, D R BOGGS, H R WARNER G E CARTWRIGHT, AND M M WINTROBE Leukokinetic Studies XIII A nonsteady state kinetic evaluation of the mechanism of cortisone induced granulocytosis *J Clin Invest* 47 249-260, 1968

BISHOP, C AND D M SURGENOR *The Red Blood Cell* New York Academic press 1964

BLEICH, H L Computer evaluation of acid base disorders *J Clin Invest* 48 1689-1696 1969

BLOOM, B R AND B BENNETT Mechanism of a reaction *in vitro* associated with delayed type hypersensitivity *Science* 53 80-82 1966.

BLOOM, B R, AND L JIMINEZ Migration inhibitory factor and the cellular basis of delayed type hypersensitivity reactions *Amer J Path* 60 453-468 1970

BLOOM W Lymphocytes and monocytes theories of hematopoiesis In *Handbook of Hematology*, edited by H Downey New York Hoeber 1938 vol I p 373-435

BOLLMAN J L J C CAIN AND J H GRINDLEY Techniques for the collection of lymph from the liver small intestine or thoracic duct of the rat *J Lab Clin Med* 33 1349-1352 1948

BOND V P T M FLIEDNER AND J O ARCHAMBAUE *Mammalian Radiation Lethality* New York Academic Press 1965

BOND V P M ODARTCHENKO H COTTIER L FEINENDEGEN AND E P CRONKITE The kinetics of the more mature erythrocytic precursors studied with tritiated thymidine in *Erythropoiesis*, edited by L O Jacobson and M Doyle New York Grune 1962 p 173-183

BORN G V R AND M J CROSS Effects of inorganic ions and of plasma proteins on the aggregation of blood platelets by adenosine diphosphate *J Physiol* 170 397-414 1964

BORSOOK H AND G L KEIGHLEY Continuing metabolism of nitrogen in animals *Proc Roy Soc (London) [Biol] Ser B* 118 488-521 1935

BOVE J R Coombs' (antiglobulin) test *J A M A* 200 459 1967

BRAIN M C J V DACIE AND D O B HOURIHANE Microangiopathic haemolytic anemia the possible role of vascular lesions in pathogenesis *Brit J Haemat* 8 358-374 1962

BRECHER G AND E P CRONKITE Post radiation parabiosis and survival in rats *Proc Soc Exp Biol Med* 77 292-294, 1951

BRECHER G, E F JAKOBEK N A SCHNEIDERMAN, G Z WILLIAMS, AND T J SCHMIDT Size distribution of erythrocytes *Ann N Y Acad Sci* 99 242-261 1962

BRECHER G, M SCHNEIDERMAN AND E P CRONKITE The reproducibility and constancy of the platelet count *Amer J Clin Path* 23 15-26 1953

BROWN, E B JR R DUBACH AND C V MOORE Studies in iron transportation and metabolism XI Critical analysis of mucosal block by large doses of inorganic iron in human subjects *J Lab Clin Med* 52 335-355 1958

BRUCE, W R, AND B E MEEKER Comparison of the sensitivity of normal hematopoietic and transplanted lymphoma colony forming cells to tritiated thymidine *J Nat Cancer Inst* 34 849-856, 1965

BURNETT F M *The Clotal Selection A Theory of Acquired Immunity* Cambridge, Mass Univ Press, 1959

CAFFREY R W W O RIEKE AND N B EVERETT Radioautographic studies of small lymphocytes in the thoracic duct of the rat *Acta Haemat (Basel)* 28 145-154, 1962

CARTWRIGHT G E J W ATHENS, AND M M WINTROBE The kinetics of granulopoiesis in normal man *Blood* 24 780-803 1964

CASTLE, W B, AND G R MINOT *Pathological Physiology and Clinical Description of the anemias*, edited by H A Christian New York Oxford Univ Press 1936

CASTLE, W B, C TOWNSEND, AND C W HEATH Observations on the etiological relationship of achylia gastrica to pernicious anemia *Amer J Med Sci* 180 305-335 1930

CHANANA A D E P CRONKITE D D JOEL L M SCHIFFER AND H SCHNAPPAUF *Transplantation* 7 459-467 1969

CHANUTIN A AND R R CURNISH Effect of organic and inorganic phosphates on the oxygen equilibrium of human erythrocytes *Arch Biochem* 121 96-102 1967

CHINARD F P G J VOSBURGH, AND T ENNS Transcapillary exchange of water and of other substances in certain organs of the dog *Amer J Physiol* 183 221-234 1955

CLARK, W M *Topics in Physical Chemistry* Ed 2 Baltimore Williams & Wilkins, 1952

CLIFF W H, AND P A NICOLL Structure and function of lymphatic vessels of the bat's wing *Quart J Exp Physiol* 55 112-121 1970

CLINE M J Leukocyte function in inflammation the ingestion killing and digestion of microorganisms *Ser Haemat* III 3-16 1970

COHN Z A AND J G HIRSCH The isolation and properties of the specific cytoplasmic granules of rabbit polymorphonuclear leukocytes *J Exp Med* 112 983-1004 1960

CONRAD M E [illegible]

CONRAD M E [illegible]

COOMBS R R A [illegible] of the mechanism [illegible] *Exp. Path.* [illegible]

COOPER P [illegible] and J H [illegible] pathogenesis of [illegible] 509–522 1965

COOPER R A AND J H JANDL [illegible] red cell lipids *J. Clin. Invest.* [illegible]

COTTIER H E P CRONKITE [illegible] SINGER AND C R SIPE Studies [illegible] extracorporeal irradiation of the [illegible] reticular organs in the cat *Blood* [illegible]

COTTIER H N ODARTCHENKO [illegible] D STONER Incorporation [illegible] in the lymphoid and plasma cell [illegible] to tetanus toxoid in mice *Ann. N. Y. Acad. Sci.* [illegible]

CREESE R M W NEIL J M LEDINGHAM AND D A SHEP The terminology of acid base regulation *Lancet* [illegible]

CRONKITE E P Enigmas underlying [illegible] proliferation *Fed. Proc.* 20 649–661 1961

CRONKITE E P Erythropoietic [illegible] *(Balt.)* 45 635–37 1964b

CRONKITE E P V P BOND T M FLIEDNER AND S A WILLMANN The use of tritiated thymidine [illegible] cell proliferation In *Ciba Foundation Symposium on Hemopoiesis 1960* edited by G E W Wolstenholme and M O Connor London Churchill 1960 p 70–92

CRONKITE E P V P BOND T M FLIEDNER AND J R RUBINI The use of tritiated thymidine in the study of DNA synthesis and cell turnover in hemopoietic tissue *Lab. Invest.* 8 263–277 1959

CRONKITE E P A D CHANANA D D JOEL K R RAI AND L M SCHIFFER Influence of extracorporeal irradiation of the blood and lymph on lymphopoiesis and immunity In *Proceedings of the Symposium of International Atomic Energy Agency Vienna 1968* p 307–326

CRONKITE E P G J JACOBS G BRECHER AND G DILLARD The hemorrhagic phase of the acute radiation syndrome due to exposure of the whole body to penetrating ionizing radiation *Amer. J. Roentg.* 67 796–803 1952

CRONKITE E P AND P C VINCENT Granulocytopoiesis *Ser. Haemat.* II 3–43 1969

CUDKOWICZ G AND J H STIMPFLING Deficient growth of C57BL marrow cells transplanted in F1 hybrid mice Association with the histocompatibility 2 locus *Immunology* 7 291–306 1964

DAVENPORT, H W *The ABC of Acid Base Chemistry* Ed 4 Chicago Univ. Chicago Press 1958

DAVEY M G *The Survival and Destruction of Human Platelets* New York S. Karger 1966

DAVID J R Delayed hypersensitivity *in vitro* Its mediation by cell free substance formed by lymphoid cell antigen interaction *Proc. Nat. Acad. Sci. U.S.A.* 56 72–77 1966

DAWSON J B The ESR in a new dress *Brit. Med. J.* 1 1697–1704 1960

DEUTSCH, H F, AND H H FUDENBERG Immunoglobulin structure and function *Advances Intern. Med.* 15 377–306 1969

DEUTSCH H F C G KASPER, AND D A WALSH Rapid method for preparation of crystalline human ceruloplasmin from Cohn Fraction IV 1 *Arch. Biochem.* 99 132–135 1962

DIETRICH, A Storungen des Kreislaufes In *Pathologische Anatomie Ein Lehrbuch fur Studierende und Arzte* edited by L Aschoff Jena Germany Fischer, 1919 p 469–534

DONALDSON, V H Blood coagulation in related plasma enzymes in inflammation *Ser. Haemat.* III 39–95 1970

DONOHUE D M R H REIFF M L HANSEN V BETSON AND C A FINCH Quantitative measurements of the erythrocytic and granulocytic cells of the marrow and blood *J. Clin. Invest.* 37 1571–1576, 1958

DRABKIN D L Spectroscopy Photometry and spectrophotometry

[illegible]

[illegible]

[illegible]

phocytes and thymocytes in the rat I The route of migration from blood to spleen and lymph nodes *J Exp Med* 127 155–168 1968
GOLUB E S AND J K SPITZNAGEL The role of lysosomes in hypersensitivity reactions tissue damage by polymorphonuclear neutrophil lysosomes *J Immun* 95 1060–1066 1966
GOOD R A P G QUIE D P WINDHORST A R PAIGE G E RODE, J WHITE J J WOLFSON AND B H HOLMES Fatal (chronic) granulomatous disease of childhood a hereditary defect of leukocyte function *Seminars Hemat* 5 215–254 1968
GORDON A S R O NERI C D SIEGEL B S DORNFEST E S HANDLER J I OBUE AND M EISLER Evidence for a circulating leukocytosis inducing factor (LIF) *Acta Haemat (Basel)* 23 323–341 1960a
GORDON A S AND E D ZANJANI Hormones and erythropoiesis In *Formation and Destruction of Blood Cells*, edited by T J Greenwalt and G A Jamieson Philadelphia Lippincott 1970 p 34–64
GOWANS J L The effect of the continuous re infusion of lymph and lymphocytes on the output of lymphocytes from the thoracic duct of unanesthetized rats *Brit J Exp Path* 38 67–78 1957
GOWANS J L The recirculation of lymphocytes from blood to lymph in the rat *J Physiol (London)* 146 54–69 1959
GOWANS J L AND E J KNIGHT The route of re circulation of lymphocytes in the rat *Proc Roy Soc (London) [Biol] Ser B* 159 257–282 1964
GOWANS J L AND J W UHR The carriage of immunological memory by small lymphocytes in the rat *J Exp Med* 124 1017–1030 1966
GRANGER G A Mechanisms of lymphocyte induced cell and tissue destruction *in vitro Amer J Path* 60 469–482 1970
GRANICK S FERRITIN IX Increase of the protein apoferritin in the gastrointestinal mucosa as a direct response to iron feeding The function of ferritin in the regulation of iron absorption *J Biol Chem* 164 737–746 1946
GRIFFITHS D A A E EDWARDS C G L GACHES A H L LONG J L W WRIGHT AND J M YOFFEY Reticulocyte generation time and rate of discharge from the marrow *Acta Haemat (Basel)* 43 13–20 1970
HAHN, P F J F ROSS W F BALE, W M BALFOUR AND G H WHIPPLE Red cell and plasma volumes (circulating and total) as determined by radio iron and by dye *J Exp Med* 75 221–232, 1942
HALL J G Effector mechanisms in immunity *Lancet* 1 25–28 1969
HALL J G , AND B MORRIS The output of cells in lymph from the popliteal lymph node of sheep *Quart J Exp Physiol* 47 360–369 1962
HALL J G B MORSE G D MORENO AND M C BESSIS The ultrastructure and function of the cells in lymph following antigenic stimulation *J Exp Med* 125 91–109, 1967
HARADIN A R R I WEED AND C F REED Changes in physical properties of stored erythrocytes relationship to survival *in vivo Transfusion* 9 229–237 1969
HARDWICKE, J AND J R SQUIRE The basis of the erythrocyte sedimentation rate *Clin Sci* 11 333–355 1952
HARPER, H A *Review of Physiological Chemistry* Los Altos, Calif Lange, 1963
HARRIS, J W The role of physical and chemical factors in the sickling phenomenon *Progr Hemat* 2 47–109, 1959
HARRIS, J W *The Red Cell* 1970 Cambridge, Mass Harvard Univ Press 1970
HARTMANN, R C AND D E JENKINS, JR The 'sugar water" test for paroxysmal nocturnal hemoglobinuria *New Eng J Med* 275 155–157 1966
HARTRIDGE H Shape of red blood corpuscles *J Physiol (London)* 53 LXXXI, 1920
HEIM J W On the chemical composition of lymph from cutaneous vessels *Amer J Physiol* 103 553–558 1933
HENDERSON, L J *Fitness of the Environment* New York Macmillan, 1913
HESS M W , H COTTIER, AND R D STONER Primary and secondary antitoxin responses in thymectomized mice *J Immun* 91 425–430, 1963
HILLMAN, R S J ADAMSON AND E BURKA Characteristics of vitamin B_{12} correction of the abnormal erythropoiesis of pernicious anemia. *Blood* 31 419–432, 1968
HIRSCH E O AND F H GARDNER The transfusion of human blood platelets With a note on the transfusion of granulocytes *J Lab Clin Med* 39 556–569 1952
HÖGLUND S Deficiency and absorption of iron in man *Acta Med Scand* (Suppl 518) 24 pp 1970
HOVIG T The ultrastructure of blood platelets in normal and abnormal states *Ser Haemat* 1 3–64 1968
HOWELL W H AND E HOLT Two New Factors in blood coagulation heparin and pro antithrombin *Amer J Physiol* 47 328–341 1918
HUDSON G AND J M YOFFEY The passage of lymphocytes through the sinusoidal endothelium of guinea pig bone marrow *Proc Roy Soc (London) [Biol] Ser B* 165 486–496 1966
HUGHES W L V P BOND G GRECHER E P CRONKITE R B PAINTER H COSTER AND F G SHERMAN Cellular proliferation in the mouse as revealed by autoradiography with tritiated thymidine *Proc Nat Acad Sci U S A* 44 476–483 1958
HUISMAN T H J AND W A SCHROEDER *Critical Rev Clin Lab Sci* 1 471–526 1970
INGRAM, V M A specific chemical difference between the globins of normal human and sickle cell anaemia haemoglobin *Nature (London)* 178 792–794 1956
INGRAM V M *The Hemoglobins in Genetics and Evolution* New York Columbia Univ Press 1963
IORIO R J A D CHANANA, E P CRONKITE AND D D JOEL Studies on lymphocytes XVI Distribution of bovine thymic lymphocytes in the spleen and lymph nodes *Cell Tissue Kinet* 3 161–173 1970
ISCOVE, N N J E TILL AND E A MCCULLOCH The proliferative state of mouse granulopoietic progenitor cells *Proc Soc Exp Biol Med* 134 33–36 1970
JACOB, H S AND J H JANDL Effects of sulfhydryl inhibition in red blood cells II Studies *in vivo J Clin Invest* 41 1514–1523 1962
JACOB H S , AND J H JANDL Increased cell membrane permeability in the pathogenesis of hereditary spherocytosis *J Clin Invest* 43 1704–1720, 1964
JACOB H S , AND M L KARNOVSKY Concomitant alterations of sodium flux and membrane phospholipid metabolism in red blood cells Studies in hereditary spherocytosis *J Clin Invest* 46 173–185 1967
JACOBS A AND D A GREENMAN Availability of food iron *Brit Med J* 1 673–676, 1969
JACOBS, A AND P M Miles Intraluminal transport of iron from stomach to small intestinal mucosa *Brit Med J* 4 778–781, 1969
JANDL J H , AND J H KATZ The plasma to cell cycle of transferrin *J Clin Invest* 42 314–326 1963
JANDL J H , R L SIMMONS, AND W B CASTLE Red cell filtration and the pathogenesis of certain hemolytic anemias *Blood* 18 133–148, 1961
JANOFF, A Mediators of tissue damage in human polymorphonuclear neutrophils *Ser Haemat* III 96–130, 1960
JAQUES L B E FIDLAR E T FELDSTED AND A G MACDONALD Silicones and blood coagulation *Canad Med Ass J* 55 26–31, 1946
JAYLE M F , AND G BOUSSIER Obtention a l'etat homogene d'une mucoproteine serique l'haptoglobine *Bull Soc Chim Biol* 36 959–969 1954
JEANNET M AND A HASSIG The role of lysophosphatides and fatty acids in hemolysis *Vox Sang* 9 113–127 1964
JOHNSON, H A Liver regeneration and the "critical mass hypothesis *Amer J Path* 57 1–15 1969
JOHNSON, S A Some problems associated with the histological study of cell proliferation kinetics *Cytologia, (Tokyo)* 26 32, 1961
JOHNSON S A R S BALBOA, B H DESSEL R W MONTO K A SIEGESMUND AND T J GREENWALT The mechanism of the endothelial supporting function of intact platelets *Exp Molec Path* 3 115–127 1964
JOHNSON, S A , D L VANHORN, H J PEDERSON, AND J MARR The functioning of platelets a review *Transfusion* 6 3–17 1966
JONXIS, J H P , AND J F DELAFRESNAYE 1959 *Abnormal Haemoglobin* Oxford Eng Blackwell 1959
JORPES, J E On the dosage of the anticoagulents, heparin and dicu

matol in the treatment of [illegible] Suppl 1490 7 pp 9 1

KARNOVSKY M L Metabolic [illegible] Pet 42 1463 1965 #62

KEITH N M L G ROWNTREE [illegible] A method for the determination of [illegible] Intern Med (Chicago) 11 547 576 1938

KELLER H U AND E SOPER Studies [illegible] nificance of normal sera for [illegible] Immunology 9 441 447 1965

KEYS A J T ANDERSON O MICKELSEN [illegible] AND F FIDANZA Diet and serum [illegible] of dietary cholesterol J Nutr 9 1 89 1936

KILLMANN S A E P CRONKITE I M FLIEDNER AND V P BOND Mitotic indices of human bone [illegible] cytologic distribution of mitosis Blood 9 674 750 1954

KILLMANN S A E P CRONKITE I M FLIEDNER V P BOND AND G BRECHER Mitotic indices of human bone [illegible] cells II The use of mitotic indices for estimating [illegible] ters of proliferation in serially connected proliferative [illegible] compartments Blood 21 141-163 1963

KOTANI M K SEIKI A YAMASHITA AND I HORII Lymphatic drainage of thymocytes to the circulation in the guinea pig Brit J Exp Path 27 511-520 1966

KOTANI M A YAMASHITA F RAI K SEIKI AND I HOPH Reutilization of DNA breakdown products from lymphocytes in lumen of intestine Blood 29 616 627 1967

KRETCHMAR A L E P McDONALD AND R D LANGE Mode of action of erythropoietin in polycythemic mice III The dose response curve In Hemopoietic Cellular Proliferation edited by F Stohlman Jr New York Grune, 1970 p 150-159

KROGH A The Anatomy and Physiology of Capillaries New Haven Conn Yale Univ Press 1922

LaCELLE P L Alteration of deformability of the erythrocyte membrane in stored blood Transfusion 9 238-243 1969

LaCELLE P L Alteration of membrane deformability in hemolytic anemias Seminars Hemat 7 355 371 1970

LAJTHA L G Stem cell kinetics In Regulation of Hematopoiesis edited by A S Gordon New York Appleton 1970 vol I p 111-131

LAJTHA L G R OLIVER AND C W GURNEY Kinetic model of a bone marrow stem cell population Brit J Haemat 8 442-460 1962

LAJTHA L G L V POZZI R SCHOFIELD AND M FOX Kinetic properties of hemopoietic stem cells Cell Tissue Kinet 2 39-49 1969

LAKI K The clotting of fibrinogen Sci Amer 206 60-66 1962

LANDSTEINER K The Specificity of Serological Reactions Springfield Ill Thomas 1936

LANDSTEINER, K, AND M W CHASE Experiments on transfer of cutaneous sensitivity to simple compounds Proc Soc Exp Biol Med 49 688-690 1942

LANDSTEINER, K, AND P LEVINE On individual differences in human blood J Exp Med 47 757-775 1928

LANDSTEINER, K AND A S WIENER An agglutinable factor in human blood recognized by immune sera for Rhesus blood Proc Soc Exp Biol Med 43 223-224 1940

LAPIN D M J LoBUE A S GORDEN, E D ZANJANI AND E F SCHULTZ Mechanisms of leukocyte production and release IX Kinetics of leukocyte release in leukocytophoresed rats Proc Soc Exp Biol Med 131 756-758 1969

LAWRENCE H S The transfer in humans of delayed skin sensitivity to streptoccal M substance and to tuberculin with disrupted leukocytes J Clin Invest 34 219-230 1955

LAWRENCE, H S, AND F T VALENTINE Transfer factor and other mediators of cellular immunity Amer J Path 60 437-450 1970

LEAK L V AND J F BURKE Fine structure of the lymphatic capillary and the adjoining connective tissue area Amer J Anat 118 785-809 1966

LEAK L V AND J F BURKE Ultrastructural studies on the lymphatic anchoring filaments J Cell Biol 36 129-149 1968

LEBLOND, C P, AND B E WALKER Renewal of cell populations Physiol Rev 36 255-276, 1956

[illegible] Hemopoiesis [illegible] AND [illegible] Hemopoietic [illegible] #

[illegible] Stem Cells W F DEBOER [illegible] #

MAKINODAN T W SHEPPARD AND S M DeMAYO AND [illegible] #

MACKINNESS [illegible] #

MacMILLAN [illegible] AND [illegible] M SCHAFER [illegible] Med 10 #

MAJNO G AND [illegible] PALADE [illegible] Biophys Biochem Cytol 11 571 605 1961

MAJNO G S M SHEA AND M LEVENTHAL Epithelial [illegible] transplantation in the rat II Adaptation of the [illegible] J Exp Med 23 554-567 1966 #

MALAWISTA S E [illegible] Rheum 13 191-197 1970 #

MALONEY M AND P M PATT Granulocyte [illegible] marrow to blood Blood 5 139-142 1950

MARCHESI V T AND E J GOWANS The migration of lymphocytes through the endothelium of venules in lymph nodes an electron microscope study Proc Roy Soc (Lond) Biol Sci 159 283-290 1964

MARGOLIS J Activation of plasma by contact with glass evidence for a common reaction which releases prekinin and initiates coagulation J Physiol (London) 144 1 22 1958

MARKHOVSKY Y D DANON AND A KATCHALSKY Agglutination by polylysine of young and old red blood cells Biochim Biophys Acta 124 174-179 1966

MAXIMOW A A AND W BLOOM Textbook of Histology Ed 5 Philadelphia Saunders 1948 p 93

MAZIA D The life history of a cell Amer Sci 44 1 52 1956

McCULLOCH E A J E TILL AND L SIMINOVITCH General factors effecting the control of hemopoiesis Cancer Cancer Conf 6 336-346 1964

McCULLOCH E A and J E TILL Cellular interactions in the control of hemopoiesis In Hemopoietic Cellular Proliferation edited by F Stohlman Jr New York Grune 1970 p 3 23

McGREGOR D D Bone marrow origin of immunologically competent lymphocytes in the rat J Exp Med 127 953 966 1968

McLEAN F C Application of the law of chemical equilibrium (Law of Mass Action) to biological problems Physiol Rev 18 495 532 1938

McMASTER P D The relative pressures within cutaneous lymphatic capillaries and the tissues J Exp Med 86 293-308 1947

MENDELSOHN M L Cell proliferation and tumor growth In Cell Proliferation edited by L F Lamerton and R J M Fry Oxford Eng Blackwell 1963 p 190-210

MERRELL M A GELLHORN AND L B FLEXNER Studies on rates of exchange of substances between the blood and extravascular fluid II The exchange of sodium in the guinea pig J Biol Chem 153 83-89 1944

MERRILL E W Rheology of blood Physiol Rev 49 863 888 1969

MESSNER H F M FLIEDNER AND E P CRONKITE Kinetics of erythropoietic cell proliferation in pernicious anemia Ser Haemat II 44-64 1969

METSCHNIKOFF E Sur la lutte des cellules de l organisme contre l invasion des microbes Ann Inst Pasteur (Paris) 1 321-336 1887

MEUWISSEN, H J, O STUTMAN AND R A GOOD Functions of the Lymphocytes Seminars Hemat 6 28-64 1969

MICHALKE W D M W HESS H RIEDWYL R D STONER AND H COTTIER Thymic lymphopoiesis and cell loss in newborn mice Blood 33 541-554 1969

MILLER J F A P The thymus yesterday today and tomorrow Lancet 2 1299-1302 1967

MILLER, J F A P A H MARSHALL R G WHITE The immu

nological significance of the thymus *Advances Immun* 2 111-162 1962
MILLER R L AND K L MELMON The related roles of histamine serotonin and bradykinin in the pathogenesis of inflammation *Ser Haemat* III 5-38 1970
MINOT G R AND W P MURPHY Treatment of pernicious anemia by a special diet *J A M A* 87 470-476 1926
MITCHELL G F AND J F A P MILLER Immunological activity of thymus and thoracic duct lymphocytes *Proc Nat Acad Sci U S A* 59 296-303 1968
MOLLISON P L *Blood Transfusion in Clinical Medicine*, Ed 4 Oxford Eng Blackwell 1967
MONKHOUSE F C AND D W CLARKE Further studies on the purification of antithrombin and heparin cofactor *Canad J Biochem* 35 373-382 1957
MOOLTEN S F L VROMAN G M S VROMAN AND B GOODMAN Role of blood platelets in thromboembolism *Arch Inter Med (Chicago)* 84 667-710 1949
MORAWITZ P Beitrage zur Kenntnis der Blutgerinnung *Beitr Chem Physiol Pathol* 5 133-141 1904
MORLEY A A A neutrophil cycle in healthy individuals *Lancet* 2 1220-1222 1966
MORLEY A A E A KING SMITH, AND F STOHLMAN JR The oscillatory nature of hemopoiesis In *Hemopoietic Cellular Proliferation* edited by F Stohlman New York Grune 1970
MORLEY A, AND F STOHLMAN JR Erythropoiesis in the dog The periodic nature of the steady state *Science* 165 1025-1027 1969
MORRIS B Migration intratissulaire des lymphocytes du mouton *Nouv Rev Franc Hemat* 8 525-534 1968
MORSE B S N J RENCRICCA AND F STOHLMAN JR The effect of hydroxyurea on differentiated marrow erythroid precursors *Proc Soc Exp Biol Med* 130 986-989 1969
NEURATH H *The Proteins* Ed 2 New York Academic Press 1965 vol 3
NORMAN A, M S SASAKI R E OTTOMAN AND A G FINGERHUT Lymphocyte lifetime in women *Science* 147 745 1965
NOWELL P C AND D A HUNGERFORD Chromosome studies on normal and leukemic human leukocytes *J Nat Cancer Inst* 25 85-93 1960
NYMAN M K GYDELL AND B NOSSLIN Haptoglobin und Erythrokinetik *Clin Chem Acta* 4 82-87 1959
OSMOND D G AND N B EVERETT Radioautographic studies of bone marrow lymphocytes in vivo and in diffusion chamber cultures *Blood* 23 1-17 1964
OWREN, P A Parahaemophilia Haemorrhagic diathesis due to absence of a previously unknown clotting factor *Lancet* 1 446-448 1947
PARROTT D M D M A B DESOUSA, AND J EAST Thymus dependent areas in the lymphoid organs of neonatally thymectomized mice *J Exp Med* 123 191-203 1966
PATT H M AND M MALONEY A model of granulocyte kinetics *Ann N Y Acad Sci* 113 515-522 1964
PAULING L, H A ITANO S J SINGER AND I C WELLS Sickle cell anemia, molecular disease *Science* 110 543-548 1949
PEDERSEN N C AND B MORRIS The role of the lymphatic system in the rejection of homografts A study of lymph from renal transplants *J Exp Med* 131 936-969 1970
PERUTZ M F The hemoglobin molecule *Proc Roy Soc (London) [Biol] Ser B* 173 1113-1140 1969
PERUTZ M F H MUIRHEAD J M COX AND L C G GOAMAN Three dimensional Fourier synthesis of horse oxyhemoglobin at 2 8 A resolution II The atomic model *Nature (London)* 219 131-139 1968
PETERS J P, AND D D VAN SLYKE *Quantitative Clinical Chemistry I* Baltimore Williams & Wilkins 1946
PIERCE, J V, AND M E WEBSTER Human plasma kallikreinins isolation and chemical studies *Biochem Biophys Res Commun* 5 353-357 1961
PITTS, R. F *Physiology of the Kidneys and Body Fluids* Chicago Year Book 1963
PITTS R F The role of ammonia production and excretion in regulation of acid-base balance *New Eng J Med* 284 32-38, 1971
POLLACK, W, J G. GORMAN AND V J FREDA 1969 Prevention of Rh hemolytic disease In *Progress in Hematology* edited by E B Broen and C V Moore New York Grune 1969 vol 6 p 121-147
PUTNAM F W Abnormal serum globulins In *The Plasma Proteins* edited by F W Putnam New York Academic Press 1960 vol II p 346-406
QUASTLER H AND F G SHERMAN Cell population kinetics in the intestinal epithelium of the mouse *Exp Cell Res* 17 420 438 1959
QUICK A J On the quantitative estimation of prothrombin *Amer J Clin Path* 15 560-566 1945
RABELLINO E S COLON H M GREY AND E R UNANUE Immunoglobulins on the surface of lymphocytes I Distribution and quantitation *J Exp Med* 132 156-167 1970
RACE R R AND R SANGER *Blood Groups in Man*, Ed 5 Oxford Eng Blackwell 1968
RAFF M C M STERNBERG AND R B TAYLOR Immunoglobulin determinants on the surface of mouse lymphoid cells *Nature (London)* 225 553-554 1970
REISSMANN, K R Studies on the mechanism of erythropoietic stimulation in parabiotic rats during hypoxia *Blood* 5 372-380 1950
REIZENSTEIN P G E P CRONKITE AND S H COHN Measurement of absorption of vitamin B_{12} by whole body gamma spectrometry *Blood* 18 95-101 1961
RELMAN A S Renal acidosis and renal excretion of acid in health and disease *Advances Intern Med* 12 295-347 1964
RICH A R Inflammation in resistance to infection *Arch Path (Chicago)* 22 228-254 1936
RICKARD K A A MORLEY, D HOWARD AND F STOHLMAN JR The *in vitro* colony forming cell and the response to neutropenia *Blood* 37 6-14 1971
RIFKIND R A Destruction of injured red cells *in vivo* *Amer J Med* 41 711-723 1966
ROBINSON S H G BRECHER I S LOURIE AND J E HALEY Leukocyte labeling in rats during and after continuous infusion of tritiated thymidine Implications for lymphocyte longevity and DNA reutilization *Blood* 26 281-295 1965
ROCHA E SILVA M W T BERALDO AND G ROSENFELD Bradykinin a hypotensive ans smooth muscle stimulating factor released from plasma globulin by snake venoms and by trypsin *Amer J Physiol* 156 261-273 1949
ROSENFIELD, R E F H ALLEN, JR S N SWISHER, AND S KOCHWA A review of Rh serology and presentation of a new terminology *Transfusion* 2 287-311 1962
RUCHTI, C H COTTIER E P CRONKITE C R JANSEN AND K R RAI Studies on lymphocytes XVII Differential lymphocyte depletion in lymphoreticular organs of the calf during continuous extracorporeal irradiation of the circulating blood *Cell Tissue Kinet* 3 301-315 1970
SAFIER S H COTTIER, E P CRONKITE C R JANSEN K R RAI AND H P WAGNER Studies on lymphocytes VI Evidence showing different generation times for cytologically different lymphoid cell lines in the thoracic duct of the calf *Blood* 30 301-310 1967
SCHILLING R F Intrinsic factor studies II The effect of gastric juice on the urinary excretion of radioactivity after the oral administration of radioactive Vitamin B_{12} *J Lab Clin Med* 42 860-866 1953
SCHOENHEIMER R *The Dynamic State of Body Constituants* Cambridge Mass Harvard Univ Press 1942
SCHOOLEY J C *The Thymus and Immunobiology, Structure Function and Role in Disease* New York Hoeber 1964
SCHMIDT P J Safe and effective transfusions *Progr Clin Path* 2 168-197 1969
SCHNAPPAUF H P AND U SCHNAPPAUF Drainege des ductus Thoracicus und Grosse der des (Leicht mobilisierbaren) Lymphozyten bei Kalbern Schafen und Hunden *Blut* 16 209-220 1968
SELWYN J G AND J V DACIE Autohemolysis and other changes resulting from the incubation *in vitro* of red cells from patients with congenital hemolytic anemia *Blood* 9 414-438 1954
SHARP J T Measurement of pH and blood gases in arterialized capillary blood *Med Clin N Amer* 53 137-145 1969
SHEMIN, D AND D RITTENBERG Life span of human red blood cell *J Biol Chem* 166 627-636 1946

SHERRY S A P FLETCHER AND N ALP [illegible] and fibrinolytic activity in man *Proc Soc Exp Biol Med* [illegible]

SIMON E R E R GIBLETT AND C A FINCH *Physiology of Human Blood* Seattle Univ Washington Press 1966

SINGER R B AND A B HASTINGS An improved clinical method for the estimation of disturbances of the acid base balance of human blood *Medicine (Balt)* 27 223–242 1948

SIPE C R A D CHANANA E P CRONKITE D D [illegible] AND L M SCHIFFER Studies on lymphopoiesis VII Size distribution of bovine thoracic duct lymphocytes *Proc Soc Exp Biol Med* 123 158–161 1966

SMITH J G N C PEDERSEN AND B MORRIS The role of the lymphatic system in inflammatory response *Ser Haemat* III 1–16 1970

SORKIN E V J STECHER AND J F BOREL Chemotaxis of leukocytes and inflammation *Ser Haemat* III 1 1–362 1970

STARLING E H On the absorption of fluids from connective tissue spaces *J Physiol (London)* 19 312–326 1896

STOHLMAN F JR Regulation of red cell production In *Formation and Destruction of Blood Cells* edited by T J Greenwalt and G A Jamieson Philadelphia Lippincott 1970 p 65–82

STOHLMAN F JR G LUCARELLI D HOWARD B MORSE AND B LEVENTHAL Regulation of erythropoiesis XVI Cytokinetic patterns in disorders of erythropoiesis *Medicine (Balt)* 43 631–660, 1964

STRUMIA M M W H CROSBY J G GIBSON T J GREEN WALT AND J R KREVANS General principles of blood transfusion *Transfusion* 4 303–346 1963

STRYCKMANS P, E P CRONKITE J FACHE T M FLIEDNER AND J RAMOS Deoxyribonucleic acid synthesis time of erythropoietic and granulopoietic cells in human beings *Nature (London)* 211 717–720 1966

SWANK R L, J G ROTH AND J JANSEN Screen filtration pressure method and adhesiveness and aggregation of blood cells *J Appl Physiol* 19 340–346, 1964

TEITEL P Disk sphere transformation and plasticity alteration of red blood cells *Nature (London)* 206 409–410 1965

TILL J E AND E A MCCULLOCH A direct measurement of the radiation sensitivity of normal mouse bone marrow cells *Radiat Res* 14 213–222 1961

TILL, J E, E A MCCULLOCH, R A PHILIPS, AND L SIMINOVITCH Analysis of differentiating clones derived from marrow *Cold Spring Harb Symp Quant Biol* 32 461–464 1967

TILL J E E A MCCULLOCH AND L SIMINOVITCH A stochastic model of stem cell proliferation based on the growth of spleen colony forming cells *Proc Nat Acad Sci USA* 51 29–36 1964

TOCANTINS L M The bleeding time *Amer J Clin Path* 6 160–171 1936

TYLER, R W AND N B EVERETT A radioautographic study of hemopoietic repopulation using irradiated parabiotic rats Relation to the stem cell problem *Blood* 28 872–890, 1966

VALERI C R Recent advances in techniques for freezing red cells *Critical Rev Clin Lab Sci* 1 381–425 1970

VEECH, R L, G A ROGENESS, AND H WEIL MALHERBE Formation of protoporphyrin from haemoglovin *in vitro Biochem J* 105 1209–1217 1967

VEJLENS G The distribution of leukocytes in the vascular stream *Acta Pathol Microbiol Scand* (Suppl 33) 241 pp 1938

VRIEDT, K E KAUDER, AND A M MAUER Defective myelopoiesis in congenital neutropenia *New Eng J Med* 283 1072–1077 1970

WADDELL W J AND T C BUTLER Calculation of intracellular pH from the distribution of 5,5-dimethyl-2 4 oxazol 1 dinedione (DMO) Application to skeletal muscle of the dog *J Clin Invest* 38 720–729 1959

WALSH [illegible] *Medicine* Philadelphia [illegible]

WASSERMAN L R AND H S MACFARLANE Diseases of lymph and [illegible] *Bases* [illegible]

WEA [illegible] AND [illegible] Physiology of human [illegible] *Proc Soc Exp Biol Med* [illegible]

WEBB [illegible] AND D D MANNOR Red cell agglutination [illegible] *Nature* [illegible]

WEED R I The [illegible] *Amer J Med* [illegible]

WEED R I AND A J BOWDLER Metabolic dependence of the [illegible] *Amer J Physiol* [illegible]

WEED R I AND C F PEED Membrane and other factors [illegible] *Amer J Med* [illegible]

WENNESTEN R S AND N S MOLNTI Ultrastructure of red cell membrane *Seminars Hematol* [illegible]

WEISS L The structure of the spleen [illegible] *Amer J Anat* [illegible]

WEISS L The structure of bone marrow Functional interrelationships of vascular and hematopoietic compartments in experimental hemolytic anemia an electron microscope study *J Morph* 117 467–537 1965

WEISS L AND M TAVASSOLI Anatomical hazards to the passage of erythrocytes through the spleen *Seminars Hemat* 7 372–380 1970

WELCKER, H *Analysen zum Beweise der Blut Flecken eines Mittels zur Erforschung des Blutarbestoffgehaltes* Giessen J Ricker 16 pp 1854

WELLS R E JR Rheology of blood in the microvasculature *New Eng J Med* 270 832–838 889–893 1964

WHANG J E FREI III J H TJIO P P CARBONE AND G BRECHER The distribution of the Philadelphia chromosome in patients with chronic myelogenous leukemia *Blood* 22 664–673 1963

WHIPPLE G H Hemoglobin regeneration as influenced by diet and other factors *Nobel Prize Lecture J A M A* 104 791–793 1935

WHITE A, P HANDLER AND E L SMITH *Principles of Biochemistry* Ed 4 New York Blakiston 1968

WIENER A S AND I B WEXLER *Heredity of the Blood Groups* Ed 1 New York Grune 1958

WINTERS R W K ENGEL, AND R B DELL *Acid Base Physiology in Medicine A Self Instruction Program* Westlake Ohio London Co 1967

WINTROBE M M *Clinical Hematology* Ed 6 Philadelphia Lea & Febiger 1967

WOODRUFF J AND B M GESNER The effect of neuramidase on the fate of transfused lymphocytes *J Exp Med* 129 551–567 1968

WRIGHT H P Changes in adhesiveness of blood platelets following parturition and surgical operations *J Path Bact* 54 461–468 1942

WU A M J E TILL L SIMINOVITCH AND E A MCCULLOCH A cytological study of the capacity for differentiation of normal hemopoietic colony forming cells *J Cell Physiol* 69 177–184 1967

WU, A M J E TILL, AND L SIMINOVITCH AND E A MCCULLOCH Cytologic evidence for a relationship between normal hematopoietic colony forming cells and cells of the lymphoid system *J Exp Med* 127 455–463 1968

YOFFEY J M AND F C COURTICE *Lymphatics, Lymph and Lymphoid Tissue* Ed 2 Cambridge Mass Harvard Univ Press 1956

ZUCKER FRANKLIN D Electron microscopic studies of human granulocytes Structural variations related to function *Seminars Hemat* 5 109–133 1968

5

The Excretion of Urine

5 URINE FORMATION

Introduction

STRUCTURE OF THE HUMAN KIDNEY

The functional unit of the kidney is the *nephron* (Fig 5 1), each kidney being composed of approximately one million such units The nephron begins at a blind end enclosing a leash of capillaries In this structure, known as the *glomerulus* (*renal corpuscle, Malpighian corpuscle*), urine formation begins by the separation, from the blood perfusing the glomerular capillaries, of an essentially protein-free filtrate of plasma The glomerular space or *Bowman's capsule* is continuous with the lumen of the *renal tubule*, a long coiled structure structurally and functionally divisible into three segments proximal convoluted tubule loop of Henle, and distal convoluted tubule The nephron as an individual unit ends with the distal convoluted tubule, but groups of distal convoluted tubules empty into a series of *collecting tubules*, which in turn coalesce to form *collecting ducts* The collecting ducts in turn empty into the short *papillary ducts of Bellini* which open into the renal calyces at the tips of the papillae

The glomerulus consists of a leash of capillaries interconnected by short anastamotic channels These capillaries are invaginated into and covered on their outer surfaces by the visceral layer of the epithelium of Bowman's capsule This layer of flattened cells is continuous with the parietal layer of epithelium and with the lining cells of the renal tubules Blood enters the glomerulus from the short afferent arteriole, and the coalescence of the glomerular capillaries forms the efferent arteriole, the glomerular capillaries are thus unique in being interposed between two arterioles. The glomerulus averages about 200 μm in diameter and the total surface of glomerular capillaries in the two human kidneys is estimated to be about 1.5 m.^{2}. The glomeruli are found only in the cortex.

The proximal convoluted tubule is a tortuous structure averaging about 55 μ in diameter and 14 mm in length It lies in the renal cortex usually in close proximity to the glomerulus to which it is conjoined It is lined with a single layer of cuboidal cells, each interdigitated with the adjoining cells in a highly complex manner The border facing the tubule lumen is characterized and distinguished from that of other segments by delicate villiform projections which constitute what is known as the brush border The terminal portion of the proximal tubule becomes straight and dips toward the medulla to become the loop of Henle

The loop of Henle commences with a descending limb which dips down into the medulla As it does so, the cuboidal cells are suddenly replaced by the flattened cells of the *thin segment* The loop penetrates into the medulla for a varying distance, the depth depending upon the location of the glomerulus from which the tubule is derived Those loops derived from tubules with glomeruli situated near the surface of the kidney penetrate only a short distance into the medulla before turning back toward the cortex, those arising from tubules with glomeruli close to the corticomedullary junction (*juxtamedullary glomeruli*) penetrate deeply to near the tip of the papilla before turning back The first part of the ascending limb does not differ in general appearance from the thin portion of the descending limb, but as the ascending limb reaches the outer medullary zone it widens and becomes lined with a cuboidal epithelium which continues into the distal tubule The tubule continues into the cortex and comes into contact with the glomerulus from which it arose, this contact being marked by a region of closely packed nuclei designated as the *macula densa*

The distal convoluted tubule may be defined as beginning at the macula densa Like the proximal tubule with which its loops are generally entwined, the distal convoluted tubule lies in the cortex It is distinguished by

tain substances from the peritubular blood and deposits them in the fluid within the lumen Thus, three individual processes, glomerular filtration, tubular reabsorption (from lumen to blood), and tubular secretion (from blood to lumen) are involved in the process of urine formation These three processes will be considered in detail in subsequent sections

Glomerular Filtration

The idea that the glomerulus is a device for filtering fluid from the blood perfusing the glomerular capillaries goes back to Bowman who in 1842 first described the relationship of the glomerular capsule to the renal tubules Bowman believed that only the water of the urine was contributed by this filtration process, the solutes being added by secretion in the tubule Shortly thereafter Ludwig (1844) offered the view, which has since been amply confirmed, that the glomerulus acts as an ultrafilter, separating from the plasma a solution which contains all of the solutes of plasma except the proteins and each at a concentration identical to that in the plasma water except as modified by the absence of protein The cornerstones of the evidence indicating that the glomerulus is indeed an ultrafilter were supplied by A N Richards and his associates who showed that the composition of fluid collected from glomeruli in amphibians was, with respect to an extensive list of constituents, that required of a plasma ultrafiltrate The studies were later extended, with respect to a limited number of substances, to fluid collected directly from mammalian glomeruli by Walker and co-workers (1941) In addition it was shown that the hydrostatic pressure in the glomerular capillaries is sufficient to supply the driving force for the filtration process

VARIABLES THAT AFFECT FILTRATION

As in the case of other capillary beds, and in accord with the Starling hypothesis, the rate of filtration through the walls of the glomerular capillaries is determined by a balance of forces favoring and opposing filtration The main driving force is the hydrostatic pressure—the intracapillary blood pressure (P_b) Opposing filtration is the osmotic pressure exerted by the plasma proteins (π_b) and the hydrostatic pressure of the fluid in Bowman's capsule (P_c) The rate of glomerular filtration (GFR) may thus be expressed as

$$GFR = K_p([P_b - P_c] - \pi_b) \quad (1)$$

where K_p is a constant relating net filtration pressure to filtration rate and thus related to the area and the permeability of the glomerular capillary walls

The glomerular capillaries differ from those of other capillary beds in being interposed between two arterioles, in maintaining a higher hydrostatic pressure, and in having a higher permeability Pappenheimer and his associates (1951) have estimated that the permeability of glomerular capillaries is approximately one hundred times that of the capillaries in the limbs (muscles), the difference is presumably largely in the fraction of the total capillary surface occupied by the pores through which filtration actually occurs The membrane which constitutes the significant filtration resistance and which is penetrated by the pores of critical size is probably the capillary basement membrane

Some idea of the size of the pores can be obtained from the characteristics of those solutes that are filtered through the glomerular capillary membrane as compared with those that are not The normal glomerular membrane is virtually completely impermeable to the proteins of plasma The smallest of these, albumin, has a molecular weight of about 70,000 and appears in the glomerular filtrate either not at all or at a very low concentration (less than 0 2 to 0 3% of its concentration in plasma) Recent measurements in samples from the early part of the dog proximal tubule indicate that the concentration is usually less than 5 mg/100 ml Hemoglobin (molecular weight 68,000), on the other hand, is excreted at a rate which would indicate that it appears in glomerular filtrate at a concentration equal to about 5% of its concentration in the plasma. (The concentration referred to is that not combined, in the plasma, with larger protein molecules.) The hemoglobin molecule is somewhat less elongated than the albumin molecule, but the difference is relatively small In any case, it is clear that the pores in the glomerular membranes cannot be far removed in size from the dimensions of these molecules

Also pertinent is the passage through the

glomerular membrane of certain polysaccharides. Most important among these is inulin, a polymer of fructose which has found a central role in the study of renal function. Its molecular weight is usually estimated to be 5200, but different inulin molecules probably differ in size, some being smaller, others larger. In addition the molecule is greatly elongated, a characteristic which reduces its diffusibility and filtrability compared to a compact molecule of the same molecular weight. Nevertheless, the glomerular membrane appears to offer no hindrance to the passage of inulin, and its concentration in glomerular filtrate is the same as in the plasma from which the filtrate is derived. Dextrans (polysaccharides composed of glucose) of a molecular weight approximately equal to that of inulin have been found to pass the glomerular membrane freely, but their passage is progressively restricted as their molecular weight increases and those with molecular weights above 55,000 do not appear in the urine from normal kidneys (i.e., kidneys producing protein-free urine) (Wallenius, 1954).

Although different theoretical treatments will yield different dimensions for the size of the pores in the glomerular membrane, the permeability can be interpreted as attributable to pores approximately 100 Å in diameter.

Equation (1) indicates that glomerular filtration is a function of three variables as well as the area and permeability of the glomerular membrane. It is important to consider the possible contribution of changes in each of these to the physiological regulation of glomerular filtration.

The Glomerular Membrane *1) Permeability.* The permeability of the glomerular membrane per unit of functioning surface is presumably fixed by its anatomical nature and not subject to physiological variation. Although changes do occur with disease, the effect of change in permeability as opposed to change in effective filtering area is difficult to evaluate. The diseased membrane may permit the passage of molecules larger than those which normally penetrate, but the predominant effect appears to be a decreased permeability.

2) *Area.* The effective area available for filtration depends upon the number of glomeruli active at any time and the number of functioning capillaries in the glomeruli. The [illegible] in the number of functioning glomeruli [illegible] are [illegible]. In the interpretation [illegible] and there are those who believe that variation in the number of glomeruli contributing to the formation of urine is a major factor in determining the rate of glomerular filtration. Subsequent treatment of the subject in this section will be based on the assumption that all the glomeruli are continuously involved in the formation of glomerular filtrate.

Capillary Blood Pressure (P_c). The major variable in the regulation of the rate of glomerular filtration is the pressure within the glomerular capillaries. The pressure at which blood enters the afferent arterioles is not much lower than that in the aorta, the decrease in pressure between aorta and glomerular capillaries is largely determined by the resistance of the afferent arterioles. Thus other factors remaining equal, constriction of the afferent arterioles will decrease and dilatation will increase glomerular filtration. The effect of changes in the caliber of the efferent vessels is in general the converse of parallel changes in the afferent vessels.

It is to be noted that it is the *pressure not the blood or plasma flow*, in the glomerular capillaries which determines glomerular filtration rate. Thus the rate of glomerular filtration need not bear any particular relationship to the cardiac output. A low level of cardiac output is, at least in theory, quite compatible with a normal rate of filtration provided this output is sufficient to maintain a normal blood pressure and provided that renal vasoconstriction is not a part of the general vasoconstriction necessary to maintain the normal arterial pressure.

Although the renal nerves play a role in the regulation of the renal arteriolar resistances and thus in the maintenance of a stable rate of glomerular filtration (and, particularly in man, this function is maintained at a remarkable constant level under normal conditions), the nerves are not essential to this stability. The kidney even when totally isolated, shows only slight changes in glomerular filtration rate (and blood flow) as perfusion pressure is varied over a wide range. This process, frequently designated by the term autoregulation, has at-

tracted considerable interest among physiologists It is probably attributable to spontaneous adjustments of the arteriolar resistances

Intracapsular Hydrostatic Pressure (P_c) Under ordinary conditions variations in the pressure within Bowman's capsule probably contribute little to changing glomerular filtration rate With ureteral obstruction this pressure may rise until the rate of filtration is reduced to negligible levels, but such obstruction does not, of course, occur under normal circumstances It is also likely that in the extremes of solute diuresis produced experimentally, the pressure required to drive the greatly increased flow of urine against the resistance offered by the tubules may be sufficient to reduce filtration

Colloid Osmotic Pressure of the Plasma Proteins (π_c) The osmotic effect of the plasma proteins is the equivalent of a hydrostatic pressure of some 25 to 30 mm Hg opposed to glomerular filtration Glomerular filtration must, therefore, cease when the net hydrostatic pressure across the capillary membrane falls to this level Although this osmotic effect is large enough for changes to exert appreciable effects on net filtration pressure, it is doubtful whether such changes contribute significantly to the physiological regulation of glomerular filtration Experimentally, it can be shown that sudden dilution of the plasma proteins by the administration of isotonic sodium chloride solution produces an increase in glomerular filtration and urine flow (dilution diuresis) but the changes are small compared to those which can be produced experimentally in other ways

MEASUREMENT OF GLOMERULAR FILTRATION RATE

The difference between the rate at which a substance is filtered at the glomeruli and the rate at which it is excreted in the urine represents the rate at which the substance is removed from or added to the urine as the latter traverses the renal tubules The rate at which a substance is filtered is given by the product of its concentration in the plasma (or more accurately, the concentration of unbound substance in the plasma water) and the rate of formation of glomerular filtrate The rate of excretion is easily determined as the product of urine flow and the concentration in the urine of the substance in question Thus the measurement of glomerular filtration rate plays a central role not only in the evaluation of glomerular filtration itself, but in the assessment of the processes of reabsorption and secretion as well

The rate of glomerular filtration can be determined if there is some substance which is neither reabsorbed nor secreted by the tubules and if the concentration of this substance in the glomerular filtrate can be deduced from its concentration in the plasma If a substance is neither reabsorbed nor secreted as the urine flows through the tubules, then, in the steady state, the rate of excretion of the substance must be the same as the rate at which it is filtered Given the rate at which the substance is filtered and its concentration in the filtrate the volume of the filtrate can be calculated

For example, consider some substance X which we will assume has the requisite characteristics of 1) being freely filterable at the glomeruli, 2) being neither reabsorbed from nor excreted into the tubule lumen Suppose we measure the flow of urine over a 10-min period during which 20 ml of urine are collected to yield a flow (V) of 2 ml /min

The concentration (U_x) of substance X in the urine is determined and found to be 120 mg /ml

The rate of excretion, then, is given by the product of flow and concentration

$$U_xV = 120 \text{ mg /ml} \times 2 \text{ ml /min} = 240 \text{ mg /min}$$

Since for substances which are neither reabsorbed nor secreted the rate of excretion is equal to the rate of filtration, the rate of filtration of substance X is also 240 mg /min But the rate of filtration of a substance is equal to the product of the rate of formation of filtrate (GFR) times the concentration of the substance in the filtrate which, in this case, is defined above as being the same as the concentration in the plasma (P_x) Assume the concentration in the plasma in this case is determined and found to be 2 mg /ml

Thus, the rate of filtration of X is the product GFR $P_x = U_xV$ and

$$\text{GFR} = \frac{U_xV}{P_x} = \frac{240 \text{ mg /min}}{2 \text{ mg /ml}} = 120 \text{ ml /min}$$

The desirability, for the evaluation of renal function, of a substance which fulfills the criteria for measuring glomerular filtration is apparent It is generally accepted that *inulin*, a starchlike polymer of fructose obtained from

dahlia tubers, has the requisite characteristics and can be used for the measurement of glomerular filtration in all vertebrates including man (Shannon, 1935, Shannon and Smith, 1935, Richards, et al , 1938)

Use of the Term Clearance The expression U_xV/P_x denotes a quantity which is generally referred to as the *clearance* of substance X and abbreviated C_x As can be seen, it is the volume of plasma needed each minute to supply the substance in question at the rate at which it is excreted in the urine It is thus a measure of the efficiency of the kidney in the excretion of a particular substance, the larger the value of its clearance, the more effectively it is excreted The term is derived from the fact that one may consider the clearance to be that volume of plasma which completely yields its content for excretion and is thus cleared of the particular substance each minute Some confusion arises when the latter concept of clearance is used since the volume cleared of a particular substance is usually virtual rather than real The same clearance is obtained if 50 ml of plasma yield all of their content, if 100 ml lose half or 200 ml yield one-fourth

The clearance of a substance such as inulin is the volume flow of glomerular filtrate By comparison of the clearance of any other substance with that of inulin we may infer the nature of the processes involved in its excretion If it is filtered but neither reabsorbed nor excreted by the tubules, its clearance must be identical with that of inulin If its clearance is lower, it must be reabsorbed from the tubules, if higher, it must be secreted by them

Thus, if P_x is the concentration of X filterable in the plasma, T_x the amount of X which is secreted into or reabsorbed from the tubule each minute and C_{In} the inulin clearance taken as equal to the rate of formation of glomerular filtrate, then,

$$C_{In}P_x = \text{filtered X}$$

$$C_{In}P_x + T_x = \text{excreted X} = U_xV$$

(where T_x will have a negative value if X is reabsorbed from the tubules)

Then

$$C_{In} + (T_x/P_x) = U_xV/P_x = C_x \qquad (2)$$

It can be seen from equation (2) that if for any reason the capacity of the tubule (T_x) to reabsorb or secrete a particular substance is limited, the clearance of that substance will approach the rate of glomerular filtration as the plasma concentration is progressively increased, since as P_x approaches infinity, the term T_x/P_x approaches zero It will be shown that this is the case with most substances transported by the renal tubules

Rate of Glomerular Filtration in Man. The clearance of inulin in a man of average size under normal basal conditions (except for the water usually administered during the determination of such clearances) averages slightly in excess of 125 ml /minute It is customary, in order to compare individuals of varying size, to normalize the clearance values by "correcting" them to what is generally considered an average normal value of body surface area, 1 73 m^2 Women have glomerular filtration rates some 10% lower than those found in men even after "correction" to the same value of surface area

The large value for the rate of formation of glomerular filtrate emphasizes the similarly great magnitude of many of the processes of reabsorption carried on by the renal tubules Thus, if we assume that the rate of 125 ml /min is maintained throughout the day we may calculate that some 7 5 liters are filtered each hour or 180 liters/day in an individual of normal size Since ordinarily such an individual will excrete only a liter or two of urine in a day, it is clear that an amount approaching 180 liters is also reabsorbed The total body water of a 70 kg man is of the order of 45 liters, so that on the average all of the body water is filtered into the renal tubules and reabsorbed four times each day On the same basis we may calculate that the extracellular fluid (approximately 12 liters) is filtered and reabsorbed some 15 times/day, the plasma volume (roughly 3 liters) some 60 times

We may further multiply this volume by its content of the various filtrable solutes of plasma to estimate the amounts of these filtered and by comparison with the amounts excreted in the urine, calculate the amounts of each reabsorbed Illustrative data are indicated in Table 5 1. It is apparent from this table that very large quantities are turned over by the kidney each day and that only a small percentage departure from normal in the process of reabsorption could result in very large changes in excretion.

Other Measures of Glomerular Filtration. In man, except for inulin, no substance has

Table 5 1
Filtration reabsorption and excretion of certain normal constituents of plasma

	Filtered/24 hr		Excreted/24 hr [a]		Reabsorbed/24 hr	
	g	*mEq or mmoles*	*g*	*mEq or mmoles*	*g*	*mEq or mmoles*
Sodium	40	24,500	3 3	150	537	24,350
Chloride	630	17,850	5 3	150	625	17,700
Bicarbonate	300	4,900	0 3	1	300	4,900
Potassium	28	700	3 9	100	24	600
Glucose	140	780	0	0	140	780
Urea	53	870	25	410	28	460
Creatinine	1 4	12	1 4	12	0	0
Uric acid	8 5	50	0 8	5	7 7	45

[a] These are typical normal values Wide variation is found depending on diet

been found which can be considered to give an accurate and reliable value for the rate of glomerular filtration Approximations can be made using one of several substances, and such approximations are useful for comparative purposes but not for the evaluation of tubular reabsorptive and secretory processes Most commonly the creatinine clearance is used and gives a value, under most conditions, not too different from the clearance of inulin

This is probably the result of compensating errors, since creatinine is excreted into the lumen by the tubules of the human kidney It is possible that it is also reabsorbed and in addition the methods usually used for determining the concentration of creatinine in plasma are not specific and include small amounts of other substances which apparently are reabsorbed Under some conditions, particularly in the presence of renal disease, the departure of the creatinine clearance from the rate of glomerular filtration may be considerable

Other substances which have been used to approximate the rate of glomerular filtration in man are certain hexitols (mannitol, sorbitol, dulcitol), particularly mannitol (Smith et al, 1940) and allantoin The mannitol clearance is about 10% lower than the clearance of inulin, indicating that about 10% of the filtered mannitol is, on the average, reabsorbed (Berger et al, 1947)

A fairly close approximation to the rate of glomerular filtration can be obtained through the use of vitamin B_{12} labeled with radioactive cobalt Although vitamin B_{12} is bound to plasma proteins, the binding capacity is relatively small and can be saturated by the administration of a moderately large dose of non-radioactive B_{12} Consequently the bound becomes a negligible fraction of the total plasma B_{12} and, since there is no measurable reabsorption or secretion of B_{12}, the clearance is close to the filtration rate (Nelp et al, 1964)

In many lower mammals, as well as in amphibians, the creatinine clearance does not have the limitations which reduce its usefulness in man The clearance of creatinine is, for practical purposes, identical with that of inulin in the dog, cat, and rabbit

Tubular Reabsorption

The quantitative importance of the process of reabsorption in the renal tubules has been indicated above by the magnitude of the amounts of material involved An understanding of the nature of the mechanisms by which this reabsorption is effected is essential to a grasp of the function of the kidney

MECHANISMS OF REABSORPTION

The processes involved in tubular reabsorption are divided into two categories, passive and active (Ussing, 1949)

Passive Reabsorption. It is clear that the volume of the urine is greatly reduced as it flows through the tubules If, in the course of this reduction in volume, some solute is retained in the lumen to a greater extent than water, its concentration will rise above that in the glomerular filtrate and above the concentration in the plasma and interstitial fluid which surround the tubules A concentration gradient for the substance is thus created between the lumen and the environment of the tubule and, provided the tubule is permeable to some extent to the substance in question, it

will diffuse from the lumen of the tubule to interstitial fluid and blood This type of process is designated by the term *passive reabsorption* and is typified by the reabsorption of urea, the excretion of which will be considered in more detail later

More generally, the movement of any substance across a biological membrane (in this case the complex membrane is the renal tubule epithelium) is considered *passive* when it can be explained entirely by gradients of concentration and/or (in the case of charged particles) gradients of electrical potential In such a context, the movement of urea is a passive process requiring no mechanism in the tubule cells specific for urea and no expenditure of energy by the tubule cells for the movement of urea itself However, it should be noted that work must be done to raise the urea concentration so that it may diffuse out passively It will be shown that in this instance the primary active process is the removal of salt from the lumen by the activity of the tubule cells, the movement of water is the passive consequence of the transport of salt and the movement of urea is in turn secondary to the movement of water

Active Reabsorption The reabsorption of many substances from the renal tubules cannot be explained by gradients of concentration or electrical potential The process involved in the movement of these solutes is designated as *active reabsorption* The best known and possibly simplest example of an active reabsorptive process is probably that for glucose A consideration of glucose reabsorption will illustrate the characteristics of these mechanisms

ACTIVE REABSORPTION THE REABSORPTION OF GLUCOSE

Under normal conditions approximately 100 mg of glucose are filtered at the glomerulus each minute Of this, none but the merest trace is excreted in the urine Clearly, the transport of glucose is effected against a *concentration gradient* and, since glucose carries no electrical charge, its transport is, by definition, *active*

If the concentration of glucose in the plasma is progressively increased, it is found that, for a time, the urine remains free of glucose However, when a sufficiently high plasma glucose concentration is reached, glucose appears in the urine in appreciable quantities Further increments in plasma glucose concentration are then accompanied by proportional increments of glucose excretion The basis for the relationship of plasma glucose concentration and glucose excretion becomes clear when the glomerular and tubular events are analyzed by the simultaneous determination of glomerular filtration rate (inulin clearance) From the product of inulin clearance and plasma glucose concentration,[1] the filtered glucose is calculated, the difference between filtered and excreted glucose is the amount reabsorbed It is then found that as the plasma glucose concentration increases (the filtered glucose of course, increasing proportionately) the reabsorbed glucose increases remaining equal to the filtered, until a limiting value of glucose reabsorption is reached As the filtered glucose increases beyond this level there is no further increase in glucose reabsorption and the excess glucose is excreted in the urine (Shannon & Fisher 1938) These relationships are illustrated in Figure 5 3 The relationship among the various quantities is expressed by the equation

$$U_G V = C_{In} P_G - T_G$$

where T represents the amount reabsorbed per minute from the tubule and the subscripts G and In refer to glucose and inulin respectively

T_G is seen to increase, remaining equal to the filtered glucose (frequently referred to as the *load* or *filtered load*), until T_G reaches a maximum This maximum, which represents the capacity of the tubules to transport glucose, is conventionally designated as the glucose T_m or T_{mG} The existence of such a limited capacity subject to saturation is a characteristic feature of active transport processes and is one of the criteria by which they may be identified

The limitation which characterizes the transport mechanism involved in glucose reabsorption and other similar mechanisms is believed to derive from the involvement of a "carrier" in the process In this view, the tubule cell membrane is virtually impermeable to glucose in the free state However, the

[1] In determining the amount filtered in the case of substances, such as glucose, which are rapidly metabolized by the tissues, it is important to measure the arterial plasma concentration, since the venous concentration may be appreciably lower and since it is arterial plasma from which glomerular filtrate is derived

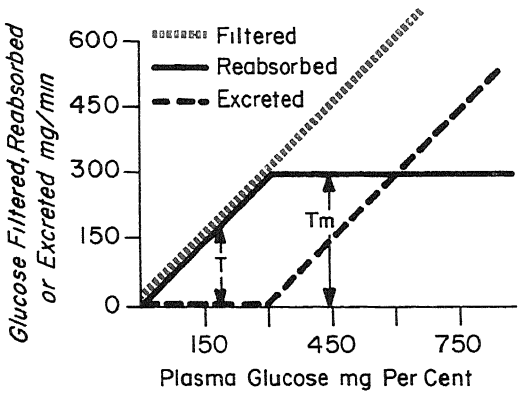


Fig 5 3 Diagramatic and idealized representation of the relationship among filtration reabsorption and excretion of glucose The figures are calculated on the assumption that glomerular filtration rate was 100 ml per minute

membrane contains certain molecules which can combine reversibly with glucose, and in this state the glucose, now in combination with its carrier, is able to penetrate the membrane Having reached the inside of the tubule cell, the carrier-glucose complex dissociates releasing free glucose which, we may assume, is free to diffuse out of the cell into the peritubular fluid and blood The carrier is then available to repeat the process

The reactions involved may be represented

$$A + B \overset{(1)}{\rightleftharpoons} AB \overset{(2)}{\rightarrow} T + B$$

where A is glucose in the tubule lumen, B the carrier, AB the glucose-carrier complex and T the glucose which has been transported (e g , into the tubule cell) Some step in the process requires an input of energy since the expenditure of energy is necessary to drive the reaction "uphill" from the low concentration in the lumen to the higher one in the cell This energy might be involved in driving reaction (2) or possibly the carrier might be modified in the transport process and require activation before it can again combine with glucose in the lumen. When the mechanism is saturated, all the carrier is interpreted as in the combined state AB and the maximum rate of transport (T_m) reflects the total amount of carrier in the tubule cells and the rate constant for the dissociation of AB to yield B and T Although, theoretically, the rate of such transport processes could be limited by the amount of energy available to drive them, there is no evidence that this actually occurs, and the same tubule cells are able simultaneously to carry out the transport of a number of solutes among which there is no mutual interference

Although the existence of carriers with the properties described is fairly generally accepted, it should be noted that none has ever been isolated or identified chemically The interaction between carrier and transported substance may be considered analogous to that between an enzyme and its substrate, in fact the reactions suggested above correspond to those to which the kinetic behavior of enzymatic reactions is attributed Transport reactions also show certain other characteristics of enzymatic reactions, particularly in being subject to inhibition, both competitive and noncompetitive

The glucose transport mechanism is fairly specifically inhibited by the glycoside, *phlorizin.* The administration of this inhibitor in adequate dosage completely abolishes glucose transport so that the clearance of glucose becomes equal to the glomerular filtration rate This property is also shared by the glycoside of the adrenocortical steroid, desoxycorticosterone Several monosaccharides (e g , xylose, fructose, and galactose) are reabsorbed to some extent apparently by the same transport mechanism responsible for the reabsorption of glucose, and their reabsorption is also inhibited by phlorizin

Micropuncture studies have clearly shown that the reabsorption of glucose occurs high in the proximal convoluted tubule and that when the plasma glucose concentration is normal the glucose has largely been removed before the urine reaches the middle of the proximal tubule (Fig 5 4)

In human beings, the normal capacity to reabsorb glucose (glucose T_m) averages approximately 375 and 300 mg /min /1 73 m ^{2} of body surface area in men and women, respectively If we compare the capacity to reabsorb glucose with the rate of glomerular filtration we may deduce what concentration of glucose in the plasma would be required to yield enough filtered glucose to saturate the glucose T_m Since the average normal rate of glomerular filtration is 125 ml /min it is apparent that a plasma glucose concentration of 3 mg /ml or 300 mg /100 ml is required to yield a filtered load equal to reabsorptive capacity of 375 mg./min. From the idealized and oversimplified diagram of excretion and reabsorption versus load plotted in Figure 5.3 it would be predicted that *no* glucose should appear in the

urine until the plasma glucose is 300 mg /100 ml , at that point, filtered load should be exactly equal to reabsorptive capacity and reabsorptive capacity should be fully saturated Further increases in plasma glucose should give no increase in glucose reabsorption In actual practice this is not the case In Figure 5 5 is reproduced the observed relationship between plasma glucose (filtered glucose) and excretion and reabsorption in normal man It is seen that glucose begins to appear in the urine when the filtered glucose is only about three-fourths of the capacity of the tubules to reabsorb and that the reabsorption does not reach its maximum rate until the amount filtered is some 50% greater than this maximum

The departure from the idealized relationship is due to two factors Two assumptions are implicit in the construction of the idealized relationship, and neither is completely valid *1*) each tubule reabsorbs all the glucose presented to it until its capacity is fully saturated, and *none* of the filtered glucose escapes reabsorption until saturation is achieved, and *2*) all of the tubules are functionally identical so that the kidney may be looked upon as a single nephron We may now consider why these assumptions are valid only as first approximations

The hypothetical basis of transport in terms of the formation of a substrate-carrier complex was presented above The initial reaction is visualized as the reversible formation of the complex

$$A + B \rightleftharpoons AB$$

If we write the mass action expression for this reaction, we obtain

$$K = [A][B]/[AB]$$

where the brackets signify concentration and K is the dissociation constant for the substrate-carrier complex The smaller the value of K, the greater the affinity of the carrier for the transported substance

It is apparent from this that the residual (unreabsorbed) substrate will be very small only *1*) if the value of K is very small or *2*) when there remains an appreciable concentration of unsaturated carrier (B) Conversely (once the total A available is equal to the total transport capacity) there will be negligible unutilized transport capacity only *1*) when K is very small or *2*) when there remains a considerable amount of unreabsorbed substrate.

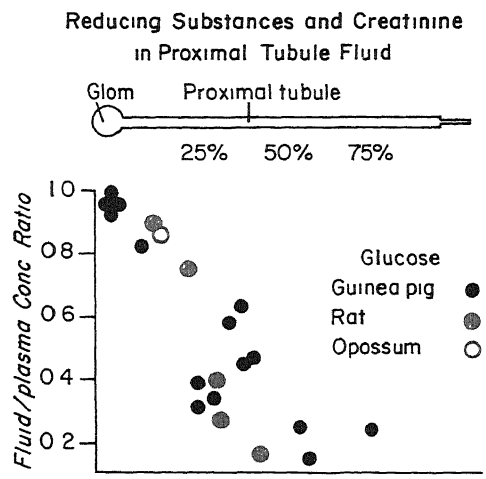


Fig 5 4 Chart showing concentration ratios of glucose between blood plasma and fluid collected from glomeruli and proximal tubules Sites of fluid collection can be identified by reference to the diagrammatic tubule (*top*) From Walker et al 1941

Thus, the relationship will approach the ideal as the value of K becomes infinitely small and for any finite value of K there will be some departure from the ideal (Fig 5 3) Actually among the various substances actively reabsorbed by the renal tubules there are varying degrees of departure from ideality presumably reflecting differences in the value of K The departure from ideality in the case of glucose is small compared to that for many other substances, suggesting a low value for the dissociation of the glucose-carrier complex

The type of relationship plotted in Figure 5 5 is commonly referred to as a *titration curve*, and the departure from the sharp break of the ideal relationship is called the *splay* of the titration curve

The nephrons are clearly not all identical nor is its reasonable to expect them to be so The heterogeneity of the nephron population which is pertinent to the shape of reabsorptive titration curves is heterogeneity of what is called *glomerulotubular* balance, that is, the relationship between the filtering capacity of the glomerulus and the reabsorptive capacity of the tubule attached to that particular glomerulus Thus if we deal with a tubule with an average capacity to reabsorb glucose but a filtration rate of twice the average, it will become saturated with glucose when the plasma glucose concentration is only half that required to saturate the average nephron Conversely nephrons with small glomerular filtering ca-

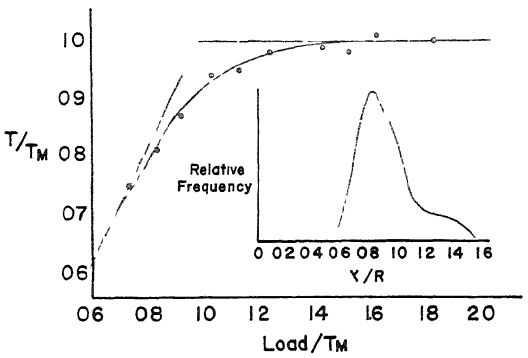


Fig 5 5 Titration of the capacity to reabsorb glucose Based on average data from 10 normal individuals The ordinate (T/T_m) is the ratio of glucose reabsorbed at a particular value of filtered glucose to the maximum capacity to reabsorb glucose in the same individual It is thus the fractional saturation of glucose reabsorptive capacity The abscissa ($load/T_m$) is the amount of glucose available for reabsorption again related to the capacity of the individual to reabsorb glucose

The frequency distribution curve shows the proportion of nephrons having various ratios (r) of filtration rate to reabsorptive capacity as related to the average of this ratio (R) for the two kidneys as a unit

The data are those of Smith Goldring Chasis Ranges and Bradley 1943

pacity and large reabsorptive capacity will become saturated only when the plasma glucose is much higher than that which would saturate the average nephron

The titration curve in Figure 5 5 can be analyzed to yield the distribution of glomerular activity in relation to glucose reabsorptive capacity if it is assumed that the titration curve in each individual nephron is ideal (that is, if the kinetic contribution to splay is disregarded). What is obtained is a distribution of the ratios of filtration to reabsorption for the nephron population as these ratios relate to the average ratio for the kidneys taken as a unit This distribution curve, included in Figure 5 5 indicates that in the normal human kidney no appreciable number of nephrons has a ratio of filtration to reabsorption of glucose departing more than 50% from the average

The earlier literature on the physiology of the kidney divides substances which may appear in the urine into *threshold* and *no-threshold substances*. A threshold substance was one which did not appear in the urine until its concentration in the blood exceeded some value, this value being designated as the renal threshold No-threshold substances were those which appeared in the urine at all concentrations in the plasma Clearly glucose (and most other substances actively reabsorbed by the tubules) would fall in the category of "threshold" substances However, with improvements in the methods of study of renal function, this terminology has in large measure lost its usefulness, and it is greatly preferable to consider the specific mechanisms involved in the excretion of a particular substance and the quantitative aspects, formerly subsumed in the "threshold," in terms of the rates of filtration, reabsorption and secretion

OTHER SUBSTANCES ACTIVELY REABSORBED BY THE TUBULES

The reabsorption of glucose is the prototype of a number of reabsorptive processes which are qualitatively similar In several instances the site of reabsorption has been identified, and in each case the proximal tubule is the segment concerned It is probable that, except for those mechanisms involved in the movements of the more important monovalent electrolytes and water (these will be considered separately later), the active reabsorptive (and secretory) processes are confined to the proximal convoluted tubules

The normal urine contains only traces of *amino acids* although, of course, a number of amino acids are present in plasma Several separate transport mechanisms are involved in their reabsorption (Pitts, 1944) The basic amino acids arginine, lysine and histidine appear to share the same mechanism since elevation of the plasma concentration of any one depresses the reabsorption of the others (Beyer et al, 1947) Reabsorptive T_m's have been demonstrated for lysine and arginine but not for histidine, the reabsorption of which is still increasing at the highest levels produced experimentally A second transport mechanism is responsible for the reabsorption of leucine and isoleucine which compete, but the capacity for reabsorption is so high that no T_m has been demonstrated Glycine appears to be reabsorbed by a separate mechanism which also is involved in the transport of proline and hydroxyproline In those instances in which T_m's have been demonstrated, there is considerable excretion of the amino acid when the reabsorptive capacity is far from saturated. Presumably this represents a relatively high degree of dissociation of the carrier-substrate

complex *Creatine* is reabsorbed by a mechanism closely related to, if not identical with, that involved in glycine reabsorption

Phosphate is reabsorbed by an active process which can be saturated on elevation of the plasma concentration The transport mechanism seems to be modified by a number of factors (Pitts and Alexander, 1944) It is probable, that parathyroid hormone depresses the phosphate transport system Acidosis increases phosphate excretion in part by what appears to be an increase in the splay of the titration curve

Sulfate reabsorption is effected by a mechanism which becomes saturated with only a slight increase in the plasma concentration above the normal level (Lotspeich, 1947) The reabsorption of sulfate can be depressed by the administration of thiosulfate which competes for reabsorption by the same mechanism

The study of *uric acid* reabsorption is complicated by the fact that uric acid is probably also secreted by the tubules (Gutman, et al, 1959) The reabsorptive capacity for urate is far greater than the amount normally contained in the glomerular filtrate and it is therefore almost certain that this capacity is never saturated in the normal individual The source of the urate normally excreted in the urine is therefore somewhat uncertain, although it might be urate which escapes reabsorption by virtue of the great splay in the reabsorptive titration curve, it is generally believed that it represents urate secreted into the tubules by a secretory process of low capacity Whatever its role may be in regulating excretion under normal conditions, urate reabsorption is remarkable for the number of substances known to inhibit it and thereby to yield marked increases in urate excretion These inhibitors include other reabsorbed compounds, such as glycine, as well as a number of drugs which are best known for their ability to inhibit the secretion of organic anions by the tubules (e.g , carinamide, probenecid, phenylbutazone)

Among other substances actively reabsorbed by the tubules are *ascorbic acid* and *betahydroxybutyric* and *acetoacetic acids*

Tubular Secretion

The processes by which materials are removed from the peritubular fluid and transferred to the tubule lumen are designated by the term *tubular secretion* The term *tubular excretion* is often used synonymously In the present section the secretion of the monovalent ions potassium, hydrogen and ammonium will be omitted, and the discussion will be concerned with the somewhat simpler transport processes involved in the secretion of certain organic electrolytes

The first substance for which tubular secretion was demonstrated was phenol red (phenolsulfonphthalein, PSP) The critical experiment was that of Marshall and Vickers (1923), who showed that as much as 75% of the phenol red contained in the arterial blood could be removed as the blood traversed the kidney, whereas, because of a high degree of binding of phenol red to plasma proteins, only 20 to 25% of the phenol red in the arterial plasma is free and filterable Clearly, no matter what fraction of the plasma was converted to glomerular filtrate, filtration alone could not account for the excretion of phenol red These experiments put an end to a prolonged controversy over the existence of tubular secretory processes

A large number of substances are now recognized to be transported into the tubules by the same mechanism responsible for the secretion of phenol red Among these are *hippuric acid* and a variety of its iodo-, hydroxy-, and amino derivatives (Smith et al, 1945), *penicillin*, 5-hydroxyindoleacetic acid (a metabolite of serotonin), a number of *sulfonphthaleins* in addition to phenol red, several of the iodinated compounds that have been used for x-ray contrast study (intravenous pyelography) of the urinary tract (Diodrast, Iopax, Neoiopax, Skiodan) [2] Also, although their tubular secretion in man has not been identified, it has been shown that glucuronides and ethereal sulfates are so transported in other animals, and the same is probably true in man It is notable that the list contains no normal constituents of the body, but under normal conditions it seems likely that the mechanism serves for the rapid excretion of the conjugates of ingested foreign substances, these conjugates being in

[2] While the iodine is important for X-ray density, it is not related to the susceptibility to transport which undoubtedly derives from the organic structures to which the iodine is attached The first three of those mentioned are derivatives of pyridine-N-acetic acid, the last is an organic sulfate

the form of hippurates, glucuronides and sulfates

The tubular secretory process, which is also located in the proximal tubule, resembles those involved in reabsorption in many respects, except for the reversal of direction

At low concentrations of the transported substance the mechanism is able to transport all of the material presented to it The most commonly studied substance, *p*-aminohippuric acid (PAH) may be used, as an example If PAH is administered intravenously in small amounts, so as to maintain a low concentration in the arterial plasma, it is found that the blood leaving the kidney in the renal vein may be almost free of PAH An average of some 90% of the PAH reaching the kidney in the plasma is removed Since some of the blood reaching the renal vein has probably perfused tissue which does not transport PAH, it is apparent that very little can be left over from that plasma which passes through the active areas This extremely effective *extraction* of PAH is the basis of the use of this substance for the estimation of renal plasma flow which is considered below

As the plasma concentration of the transported substance is increased, the rate of tubular excretion increases until a maximum is reached and no further augmentation of secretion occurs The relationship between plasma PAH concentration and PAH excretion in the urine is illustrated in Figure 5 6 The dotted line passing through the intercept is the amount of PAH filtered at the glomeruli calculated from the product of inulin clearance and the plasma concentration of free PAH (Since PAH is, to a small extent, bound to plasma albumin the total concentration must be corrected for this binding to obtain the filtered amount) At every level of plasma PAH the rate of excretion exceeds the rate of filtration

$$U_{PAH}V = C_{In}P_{PAH}f + T_{PAH}$$

where f is the fraction of the plasma PAH which is filterable. As the plasma PAH is increased, the rate of excretion at first increases sharply, the slope of the line relating excretion to plasma concentration being equal to the renal plasma flow. When the plasma concentration is further increased, the rate of excretion increases less rapidly, and it is found that the excess of excreted over filtered has become constant This excess represents the maximum

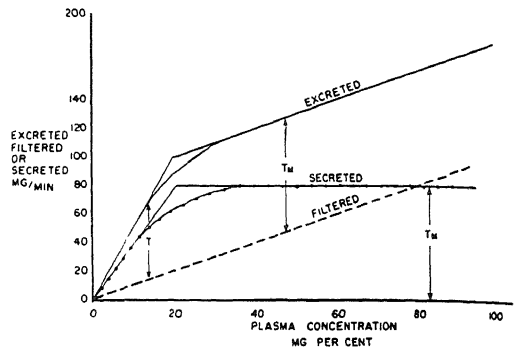


Fig 5 6 Relationship between plasma concentration and filtered secreted and excreted amounts of a substance excreted by tubule secretion See text for details

rate of tubular secretion and is designated as T_{mPAH}

If the rate of tubular secretion is determined by subtracting the filtered from the total excretion the relationship shown in Figure 5 6 is obtained Again, the actually observed titration curve differs from the idealized intersection of two straight lines and for reasons analogous to those involved in tubular reabsorption *1*) the dissociation of substrate-carrier complex and *2*) heterogeneity of nephron function The pertinent heterogeneity in this case, however, is in the ratio of perfusion to transport capacity since it is blood supply, rather than glomerular filtration, which, along with plasma concentrations, will determine the amount available for secretion by the tubules

Any of the various substances listed above as being secreted by this particular transport mechanism can be shown to compete with the others Thus a high concentration of one will reduce the secretion of another and vice versa Presumably they share a common carrier, and when some of the carrier sites are occupied by one, it has less capacity for the transport of another

Several substances that are effective inhibitors of this mechanism have been discovered (carinamide, probenecid, phenylbutazone) These compounds are competitive inhibitors but are themselves only slowly excreted because subsequent to their secretion into the tubule they are reabsorbed by the process known as *non-ionic diffusion* (see p 5–37) (It is interesting to note that these inhibitors have found their major practical application as inhibitors of the *reabsorption* of urate in the treatment of gout)

The mechanism is also inhibited by dinitrophenol and by mercurial diuretics The effect of dinitrophenol has been attributed to the property of this substance of uncoupling phosphorylation from oxidation and has been interpreted as indicating that phosphate bond energy is important in the transport process However, the finding that the dinitrophenol inhibition appears to be competitive casts doubt on this interpretation The effect of mercurial diuretics is noncompetitive and, for unknown reasons, is striking in man but entirely absent in the dog

The transport of PAH is enhanced by acetate and lactate and inhibited by a number of the dicarboxylic acid intermediates of the Krebs cycle

All of the substrates of the secretory process responsible for the tubular excretion of PAH are weak organic anions A separate and distinct mechanism is capable of transporting a variety of organic bases including N-methylnicotinamide, tetraethylammonium ion and tolazoline These compete with each other but show no interaction with the organic anions The normal substrate for this mechanism has not been recognized (Weiner and Mudge, 1964)

UTILIZATION OF TUBULAR SECRETION FOR THE MEASUREMENT OF RENAL BLOOD FLOW

As was noted above, several substances that undergo tubular excretion are very nearly completely removed from the blood that perfuses the kidney This has made them singularly useful for the estimation of renal blood flow Diodrast and *p*-aminohippurate (PAH), because they are well adapted to chemical analysis, have particularly been used for this purpose, and we will consider the use of PAH as an example

In accordance with the Fick principle, if the rate of removal of a substance from the blood as it flows through an organ is known, and if the concentration in the blood entering and leaving the organ can be determined, the blood flow can be calculated:

$$F = R/(A - V)$$

where F is the flow, R the rate of removal, and A and V the concentration in arterial and venous blood, respectively.

In applying this to the measurement of renal blood flow using PAH, the following experimental facts are important *1)* there is no metabolism, production nor storage of PAH in the kidney, all of the PAH removed from the blood is excreted in the urine, *2)* all of the PAH removed from the blood is removed from the plasma, the concentration in the red blood cells does not change significantly as blood passes through the kidney Then the above equation can be modified to yield

$$RPF = U_{PAH}V/(P^{A}_{PAH} - P^{RV}_{PAH})$$

where RPF represents renal plasma flow and P^{A}_{PAH} and P^{RV}_{PAH} represent concentrations of PAH in arterial and renal venous plasma, respectively

Now, as it stands in this form the determination of renal plasma flow would require measurement of the rate of excretion of PAH and the concentrations in arterial and renal venous blood plasma All of these are feasible in man, but the collection of renal venous blood is considerably more difficult than the remainder of the procedures However, since we know that at low concentrations of PAH in the arterial plasma the removal of PAH as blood passes through the kidney is very nearly complete, we may as a first approximation consider the concentration in renal venous plasma to be zero and write

$$ERPF = U_{PAH}V/P^{A}_{PAH} = C_{PAH}$$

or that the *effective renal plasma flow* is equal to the PAH clearance In practice it is not necessary to measure arterial plasma PAH since, when arterial concentration is constant, there is no measurable removal of PAH by peripheral tissues, and the PAH in peripheral venous plasma is the same as that in arterial plasma Therefore, the effective renal plasma flow can be simply estimated by measurement of the plasma PAH clearance when the plasma PAH is maintained at a low level (As plasma PAH is increased, the extraction becomes less complete and the clearance falls further below the true plasma flow) From the effective renal plasma flow the corresponding renal blood flow can be calculated if the hematocrit is known

$$ERBF = ERPF/(1 - Hct)$$

It is interesting to note that Homer Smith and his associates (1938) deduced that the

clearance of diodrast could be used to measure renal plasma flow in man before the technique for obtaining renal venous blood by venous catheterization had been devised and the virtually total extraction of Diodrast confirmed [3] This was based on the very high value (approximately 650 ml /min) of the Diodrast clearance, thus, plasma flow of at least 650 ml /min was required to supply the Diodrast at the rate it was excreted It was reasoned that this was already such a large volume flow that the true plasma flow could not be much higher

RENAL BLOOD FLOW AND FILTRATION FRACTION IN MAN

The remarkably high value of the renal plasma flow has been mentioned above A plasma flow of 650 ml /min corresponds to a blood flow of about 1200 ml /min , an amount approximately equal to one-fourth of the cardiac output under basal conditions The ratio of the renal blood flow to cardiac output is frequently referred to as the *renal fraction* of the cardiac output

The demonstrated equivalence of PAH clearance and (effective) renal plasma flow has made it possible to measure the renal blood flow in man with a minimum of manipulation and disturbance Thus, it has been possible to measure changes in blood flow with disease as well as to determine the effects of various changes in physiological state and the effects of drugs, etc (Chasis et al , 1938)

In considering changes in the renal plasma and blood flows, it is pertinent to take account of the fraction of the plasma lost as glomerular filtrate as the plasma flows through the glomerular capillaries Since all of the renal blood flow under normal conditions passes through the glomeruli, the volume given up as glomerular filtrate is derived from the total plasma flow and the ratio of filtration rate to plasma flow (inulin clearance to PAH clearance) is a measure of the extent to which, in the average glomerulus, the plasma volume is reduced and the plasma protein concentration increased by the process of glomerular filtration The ratio of filtration rate to plasma flow is known as the *filtration fraction*

While, as was noted in an earlier section, the rate of glomerular filtration is a remarkably stable function, the renal blood flow manifests a fairly marked degree of lability, and may be modified by any of a number of stimuli Most of the changes are mediated via the sympathetic nerve supply and are in the direction of decreases In fact, although the subject has long been debated and no consensus has been achieved, the only physiological effects clearly shown to be mediated by the nerve supply to the kidneys are vasomotor, and the renal nerves have not been shown to play a significant role in the regulation of transport processes in the renal tubules

The renal blood flow is diminished by fright, pain, exercise, epinephrine, norepinephrine and stimulation of several areas of the central nervous system, and by a number of other stimuli In general, anything that lowers blood pressure will reduce renal blood flow, not so much because of the reduction in perfusion pressure, but because the kidney takes part in the general vasoconstriction which is the homeostatic response to a lowering of blood pressure Thus there may be a marked decrease in renal blood flow with only a trivial reduction in aortic blood pressure Drugs that reduce blood pressure also reduce the renal blood flow However, in general, the reduction in flow is less than with equivalent reductions in blood pressure due to, for instance, hemorrhage or orthostasis, since the blood pressure reducing drugs usually inhibit renal as well as peripheral vasoconstriction A special case appears to be hydrazinophthalazine which may increase renal blood flow while reducing systemic blood pressure, probably through a specific renal vasodilator action

A more or less unique phenomenon is the marked renal vasodilation and hyperemia produced by the pyrogenic reaction The latter phenomenon, the most striking clinical feature of which is chills and fever, is most often a response to bacterial products (certain bacterial polysaccharides, typhoid vaccine) It is accompanied by intense hyperemia of the kidney, and the renal blood flow may reach twice its normally very high value Interestingly, the

[3] It should also be noted that if, for some reason, renal venous blood is collected so that true plasma and blood flow can be determined, PAH and similar substances lose their unique value Then any substance that is removed by, but not produced, metabolized, or stored in the kidney can be used for the measurement, since the Fick principle can be applied directly. A high degree of extraction, under these conditions, has only the virtue of minimizing experimental error.

renal response to pyrogenic materials is not abolished by the administration of antipyretic drugs although the latter may completely suppress the chills and fever

With all of the above changes in renal blood flow there are frequently only slight changes in glomerular filtration rate and such changes are frequently entirely absent Thus the changes in filtration fraction are the inverse of the changes in blood flow

What Happens to Typical Constituents

REABSORPTION AND EXCRETION OF WATER

One of the most striking features of the function of the mammalian kidney is the capacity to excrete urine differing in osmotic pressure from the blood from which it is derived It is through this capacity to excrete concentrated or dilute urine that the solute concentration (osmotic pressure) of the body fluids is maintained within remarkably narrow limits despite wide variation in the intake of solute and water When, because of water loss or solute gain, the solute concentration of body fluids rises, the urine becomes concentrated, solute is excreted with less than the equivalent amount of water, and the solute concentration of the body fluids is restored to normal Conversely, when the body fluids become diluted by water intake or solute loss, a dilute urine containing an excess of water is excreted, and the osmotic pressure of body fluids is returned to normal Before considering the mechanisms by which the renal tubules effect these changes, it may be helpful to digress briefly to review the concept of osmotic pressure and its measurement

If two solutions containing different concentrations of solute are separated by a membrane permeable to the solvent (water in all cases with which we are concerned) but not the solute, there will be a movement of water from the more dilute to the more concentrated of the solutions The movement of water is due to the fact that the presence of solute reduces the chemical activity of water, and water will move from the solution in which its activity (analogous to concentration) is high to one where it is lower Water will move from dilute to concentrated solution until both have the same solute concentration or, if the movement of water is opposed by the application of a hydrostatic pressure, until the hydrostatic pressure balances the concentration difference The hydrostatic pressure which would be required to maintain a solution at constant volume when it is separated by an ideal semipermeable membrane[4] from pure water is the osmotic pressure of that solution The term osmotic pressure is applied although the osmotic pressure is never actually determined Instead, some other property of the solution dependent upon total solute concentration is measured most commonly the freezing point, occasionally the vapor pressure

The *colligative properties* (osmotic pressure, freezing point depression, vapor pressure depression) depend on the ratio of the total number of solute particles to the number of moles of solvent The nature of the solute is, to a large extent, indifferent, ideally, a protein molecule of molecular weight 100,000 has an effect equal to that of a sodium ion of weight 23, and a mole of sodium chloride, since it dissociates to yield two particles, has twice the effect of a mole of urea or glucose One mole of ideal solute dissolved in a kilogram of water depresses the freezing point of the solution by 1 86 C Such a solution is defined as an *osmolal* solution and is said to contain one osmol/kg of water Since most biological fluids are considerably more dilute, it is convenient, in order to avoid decimals, to deal with *milliosmols* or *micro-osmols*, one milliosmol being one one-thousandth of an osmol, one micro-osmol one-thousandth of a milliosmol The milliosmolality of a solution is defined as its freezing point depression in degrees Centigrade divided by 0 00186 Thus normal plasma is about 285 milliosmolal (abbreviated 285 mOs) and has a freezing point of about −0 53°C [5]

It will facilitate consideration of the processes that take place between the filtration of

---

[4] An ideal semipermeable membrane is one permeable to solvent but impermeable to all solute

[5] Note that osmotic pressure depends on the number of solute particles per unit of solvent The amount of water in a given *volume of solution* will vary with concentration as well as with the nature of the solute There is, therefore, no simple way to convert from osmolality to a number of solute particles per unit volume of solution It is, however, the practice to disregard the inaccuracy involved and multiply volume by osmolality to obtain the approximate amount of solute in a given volume.

an isotonic fluid (one which has the same osmotic pressure as blood plasma) and the excretion of dilute and concentrated urine, if we first consider the gross behavior of water excretion and its extrarenal control (Findley and White, 1937) Under basal conditions the urine is usually moderately concentrated and the urine flow is small If under such conditions an individual ingests a moderately large volume of water, e g , one liter, it is found that after an interval of some 15 to 20 min the urine flow begins to rise The flow increases rapidly to reach a peak of some 12 to 15 ml /min at 40 min to 1 hr , and, when an amount approximately equal to that ingested has been excreted, returns to the same low level as obtained before the water ingestion (Of course, if the water ingestion is continued so that a positive balance is maintained, the high rate of flow may continue for prolonged periods) If these events are analyzed more carefully, it is found *1*) that the osmotic pressure of the blood shows a distinct drop (though only a few percent) well before the onset of the increase in urine flow (*diuresis*), *2*) that the change in urine flow is almost entirely due to a change in the amount of water excreted, manifested by a shift in urine osmolality from hypertonic to markedly hypotonic to plasma, solute excretion shows little or no change, *3*) if glomerular filtration has been measured it is found that the increase in urine flow occurs without or with only trivial changes in the rate of formation of glomerular filtrate, the change in water excretion is due to a change in the absorption of water in the renal tubules, *4*) if at the peak of the urine flow, additional water is ingested, the urine flow does not increase further but remains at a value of some 12 to 15% of the volume of the glomerular filtrate, *5*) if, at any point in the course of the diuresis, vasopressin, the antidiuretic hormone of the posterior pituitary, is administered, the diuresis is immediately interrupted, and the urine again becomes more concentrated than the blood

The concentration of the urine is normally continuously regulated by the secretion of antidiuretic hormone (ADH) by the posterior lobe of the pituitary gland This hormone is believed to be produced in the supraoptic nuclei of the hypothalamus and to migrate via the supraopticohypophyseal tracts to the posterior lobe of the pituitary where it is stored and released into the blood in response to stimuli mediated from the hypothalamus via the same tract system Within the supraoptic nuclei certain cells (*osmoreceptors*) are believed to behave as tiny osmometers, taking up water and swelling when the osmotic pressure of the surrounding body fluids is reduced and giving up water as the osmotic pressure of the bathing fluid rises The swelling and shrinking of these cells is believed to lead to the transmission of nervous impulses which modulate the release of antidiuretic hormone from the posterior pituitary Thus a reduction of the osmotic pressure of the blood leads to inhibition of ADH release,[6] an increase of osmotic pressure stimulates ADH secretion In turn, the ADH is carried by the blood to the kidney, and in the kidney ADH regulates the reabsorption and excretion of water

The work of Verney (1947), to whom we owe much of our understanding of the pituitary control of water excretion, showed that injection into the internal carotid artery of the dog of enough sodium chloride to raise the osmotic pressure of the blood by about 2% resulted in immediate inhibition of water excretion Injection of the same amount into a peripheral vessel had no such effect since it was then diluted with a large volume of blood before being delivered to the sensitive area in the brain Injection of the osmotically equivalent amount of urea into the internal carotid was also virtually without effect In this case the ineffectiveness is presumably due to the fact that urea readily penetrates cells When the urea concentration of their environment is raised, the osmoreceptor cells take up urea rather than lose water Since the receptors do not shrink, there is no induction of ADH secretion Thus, *only those solutes to which cells are relatively impermeable make up the effective osmotic pressure in response to which ADH secretion is regulated* [7]

Our knowledge of the location in the renal tubule and the nature of the processes by which water excretion is regulated is due

[6] The latent period from water ingestion to onset of diuresis is attributable *1*) to the time required for water absorption to produce sufficient dilution of the body fluids and thus inhibit ADH secretion, *2*) for the circulating ADH to disappear

[7] The above describes the normal control of ADH secretion A number of other stimuli such as fear, trauma, nausea, certain drugs, and many others may produce secretion of ADH under conditions unrelated to the osmotic pressure of the body fluids

largely to the application of the micropuncture technique—in amphibians by Richards and his associates (1936, 1938), and in mammals by Walker et al (1941), and more recently and extensively by Wirz (1956) Gottschalk (1959, 1964) and others

The glomerular filtrate has the same osmotic pressure as the plasma except for the absence of the plasma proteins [8] As the fluid flows through the proximal tubule, its volume is reduced to a fraction of that which enters the glomerular space Although, because of inaccessibility from the kidney surface no samples have been collected from the last 40% of the proximal tubule, it is generally estimated that approximately one-third of the filtered volume reaches the end of the proximal tubule and enters the loop of Henle This reduction of volume occurs *without change in osmotic pressure* whether the urine being produced is dilute or concentrated Thus the large fraction of fluid reabsorbed in the proximal tubule does not take part in the concentration and dilution processes, and, in fact, is lost to the process of osmoregulation The water thus reabsorbed in the proximal segment has been termed by Smith (1952) *obligatory water reabsorption* to distinguish it from the *facultative water reabsorption* in the more distal portions of the nephron where water reabsorption is subject to regulation in accordance with the excess or deficit of water in the body fluids

A considerable body of evidence indicates that the water reabsorption in the proximal tubule is entirely a passive process and that the active process which underlies it is the transport of salt (predominantly sodium salts) out of the tubule lumen (Windhager et al, 1959) As sodium salts are transferred from lumen to peritubular fluid, there is a tendency to produce dilution of the tubule contents and hypertonicity of the surroundings The proximal tubule is highly permeable to water and, in response to the osmotic gradient, water follows the salt out of the tubule Within the limits of the determinations the proximal tubule contents are always isotonic with the blood

Although the loop of Henle is far less accessible to puncture in situ than the convoluted tubules, it has been possible through puncture of limited segments in vivo and by perfusion in vitro to define the characteristics of the loop The descending limb has a high permeability to water so that the fluid in the descending limb retains the same osmotic pressure as the surrounding interstitial fluid Near the bend of the loop there is an abrupt decrease in the permeability to water so that the continued removal of salt leaves the fluid in the lumen dilute (Morgan and Berliner, 1968) The urine emerges into the distal convoluted tubule with its osmotic pressure reduced to about half that of plasma whether the final urine is to be dilute or hypertonic

The changes that take place in the distal tubule depend upon whether a dilute or concentrated urine is to be produced In the *formation of dilute urine*, the dilute fluid that enters the distal tubule is further diluted by the continued removal of salt, and the salt removal probably continues through the collecting system The urine is thus rendered hypotonic by the active reabsorption of salt in a region which clearly has a low permeability to water (Otherwise the osmotic gradient would be dissipated, as it is in the proximal tubule, by the escape of water to the more concentrated surroundings) The low permeability to water which characterizes the distal convoluted tubule and collecting system during the formation of dilute urine is dependent upon the absence of ADH

In the *formation of concentrated urine*, the events are similar up to the point at which the fluid enters the distal convoluted tubule In the presence of antidiuretic hormone, however, the collecting tubules and collecting ducts and possibly the distal convoluted tubules become permeable to water [9] Consequently, as the

[8] It is important to make a clear distinction between the total osmotic pressure, with which we are concerned in the process of urine dilution and concentration, and the colloid osmotic pressure or oncotic pressure due to the plasma proteins which play such an important role in the distribution of fluid across capillary membranes (because the latter are so highly permeable to all the other solutes) The colloid osmotic pressure of plasma is equivalent to about 25 mm Hg (1 to 2 mOs) while the total osmotic pressure is of the order of 6 atm (285 mOs)

[9] Antidiuretic hormone has been shown to have the property of increasing water permeability in isolated membranes (frog skin, toad skin, toad bladder) by causing the dilation of water-filled pores (Andersen and Ussing, 1957, Sawyer, 1957) It has also been shown directly, in vitro, to increase the permeability to water of rabbit cortical collecting tubules and rat medullary collecting ducts (Grantham and Burg, 1966, Morgan and Berliner, 1968) In the rat,

fluid flows through the distal nephron the osmotic gradient established by the reabsorption of sodium and chloride, is dissipated, and, despite additional salt removal in the distal nephron, the dilute character of the urine is lost so that it regains isotonicity before it enters the medullary portion of the collecting system It is important to emphasize that it enters the medulla isotonic, never hypertonic The additional water loss which renders the urine more concentrated than blood occurs in the medullary segments of the collecting system

The nature of the mechanism by which urine is rendered hypertonic to the body fluids was long one of the more intriguing problems of renal physiology, and its apparent elucidation, largely through the efforts of Wirz and his associates (1951) is one of the more important and interesting achievements Although some of the details of the mechanism are relatively complex, the essentials are quite simple and directly related to the anatomical organization of the kidney The convoluted tubules are cortical structures, the loops of Henle and the collecting ducts lie in the medulla The latter structures are thus relatively isolated from the cortical structures in which most of the transport activity of the kidney takes place In addition, the blood supply of medulla and papilla forms a looplike system derived from the efferent vessels of the juxtamedullary glomeruli, running down into the medulla, and looping back to the cortex Thus, both loop of Henle and the blood supply of the medullary region form *countercurrent* systems in which the outflowing fluid flows counter to and in proximity to inflowing fluid We will for the moment defer consideration of the importance of this particular arrangement

As was noted above, as fluid flows through the ascending limb of the loop, sodium and chloride are removed while the water remains behind in the tubule lumen and the urine becomes diluted The sodium and chloride removed are transferred to the interstitial fluid of the medulla, increasing the osmotic pressure of this fluid The diluted urine emerges into the cortex where, in the presence of ADH, it loses its excess of water and is further reduced in volume as solute is removed in the distal tubule This water loss occurs in the cortex, spatially separated from the salt left behind in the medulla, and in an area where, because of the very large flow of blood, the excess water produces only a minimal change in osmotic pressure Now the remaining urine, greatly reduced in volume and isotonic with the blood, enters the collecting system and flows through the medulla The interstitial fluid of the medulla, however, has a high osmotic pressure because of salt transported by the loop and, the collecting duct being permeable to water in the presence of ADH, the urine loses water to attain the same osmotic pressure as the interstitial fluid Thus the hypertonicity of the urine is produced These events are illustrated diagramatically in Figure 5 7 Because the volume flow through the loop is several times larger than that in the collecting ducts, a relatively small change in concentration in the loop is sufficient to yield enough solute to raise the concentration of the urine in the collecting ducts severalfold

If the sodium salts transported into the medullary interstitial fluid are to raise the osmotic pressure of the latter, they must not be immediately carried away by the blood flowing through the area The retention of excess solute in the medullary interstitial fluid is made possible by the arrangement of the blood flow which constitutes a countercurrent exchanger The principle of the countercurrent exchanger is illustrated in Figure 5 8 Outflowing blood tends to lose its solute to the inflowing blood As a result, solute is recirculated in the medulla This characteristic makes it possible not only to retain sodium salts but to trap urea as well, the special role of urea in the concentration process is considered with the excretion of urea

The loop arrangement has an additional physiological advantage As the fluid flows down the descending limb of the loop of Henle into an environment made hypertonic by the activity of the ascending limb, it loses water so as to retain the same osmotic pressure as the environment The loss of water raises the concentration of sodium salts in the loop contents so that the transport of sodium salts out of the ascending limb is against only a small gradient

---

the permeability of the distal convoluted tubule to water is regulated by antidiuretic hormone (Wirz, 1956) In the dog and in the monkey the distal tubule is unaffected by antidiuretic hormone and the permeability change is presumably first manifested in the cortical collecting tubules (Clapp and Robinson, 1966; Bennett et al , 1968)

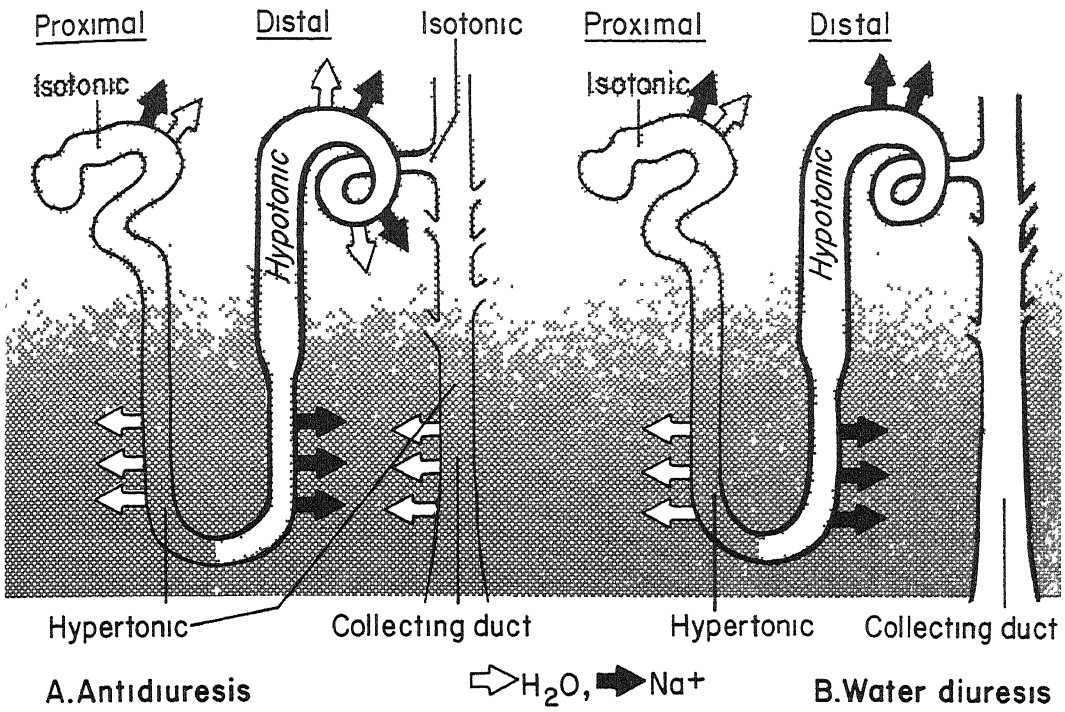


Fig 5 7 Diagramatic representation of changes in the tonicity of fluid in the tubules during the formation of concentrated and dilute urine Dark and light *arrows* indicate movements of sodium chloride and water respectively and the relative permeability of the tubule wall to water is indicated by its thickness

of concentration despite the very high concentration of sodium salts in the interstitial fluid surrounding the loop This arrangement is that of a *countercurrent multiplier*, so called because the small concentration gradient from the loop contents to environment is multiplied in the longitudinal direction from cortex to tip of papilla

The highest concentration achieved by the human kidney is generally about 1200 to 1400 mOs (a relatively modest maximum compared with some other mammals, e g , dog 2300 to 2500 mOs , rat 3000 to 3200 mOs , kangaroo rat over 5000 mOs), the lowest concentrations observed are of the order of 30 to 40 mOs Thus, the kidney is able to produce urine four to five times more concentrated than or one-seventh to one-tenth as concentrated as the blood (roughly 300 mOs)

In clinical practice, the specific gravity, rather than the osmolality, is generally used as an index of concentration and dilution Urine of specific gravity 1 008 to 1 010 is approximately isotonic with plasma, but the two scales are not directly interchangeable since the specific gravity depends not only on the concentration of solute but also on the latter's nature The chief value of the specific gravity lies in the simplicity of its determination

The highest and lowest concentrations are produced only when solute excretion is at a relatively low level, as solute excretion increases dilute urines become less dilute and concentrated urines less concentrated The basis for this change can be deduced from the mechanisms for dilution and concentration described above and is more easily understood if the urine flow is considered to be divided into two parts as suggested by Smith One fraction is considered to contain all of the solute excreted at the same concentration as plasma, the other fraction is the amount of water contained in excess of this isotonic fraction The isotonic fraction is calculated by dividing total solute excretion by the osmolality of plasma Solute excretion is given by the product of urine flow and urine osmolality $U_{os}V$ and when this is divided by the plasma

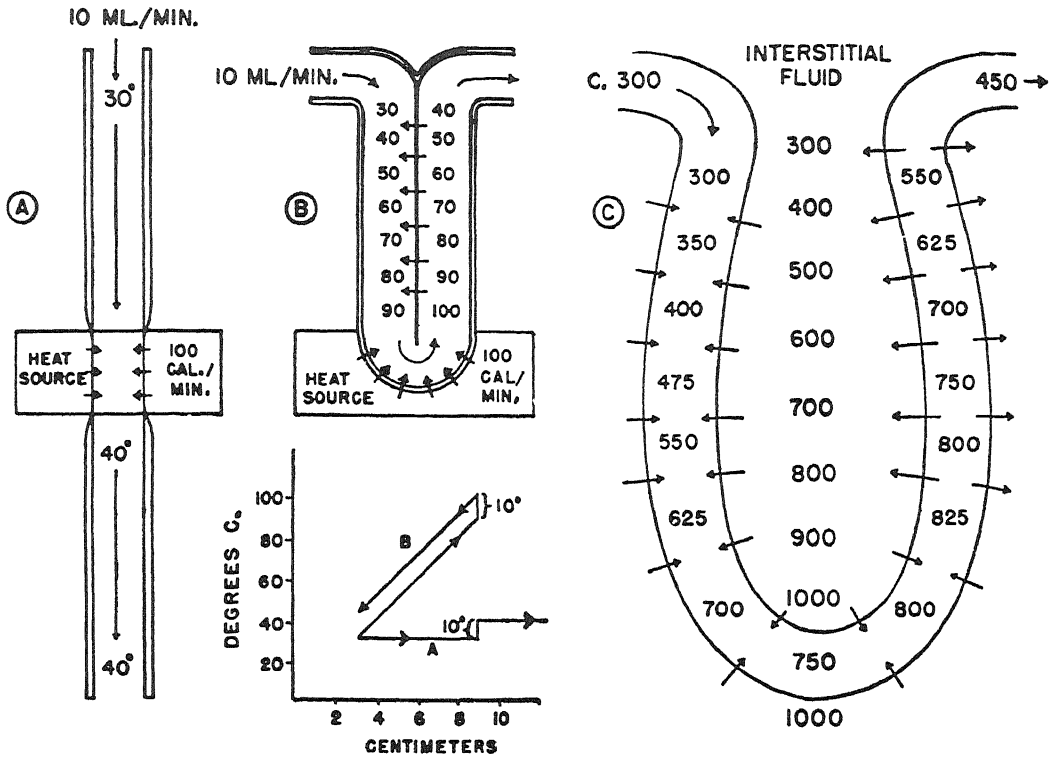


Fig 5 8 Principle of the countercurrent exchanger A and B 100 calories/min raise the temperature of 10 ml/min of water 10° in both A (straight flow) and B (countercurrent flow) However since the incoming water is heated by the outgoing water in B the *maximum temperature* attained in the countercurrent system is considerably higher than with straight flow The graph compares the temperature along the flow tubes in each system C countercurrent flow as applied to the capillary loop showing that it is not necessary for the limbs of the loop to be in direct contact In the *hypothetical* illustration given both limbs are in contact with the same interstitial fluid of progressively increasing concentration Sodium salts (*arrows*) at first enter the capillary blood later partly return to the interstitial fluid Note the analogy between B in which heat is recirculated and C in which sodium salts are similarly retained in an area

osmolality, the resulting value has the dimensions of a clearance and is generally called the osmolal clearance (C_{os})

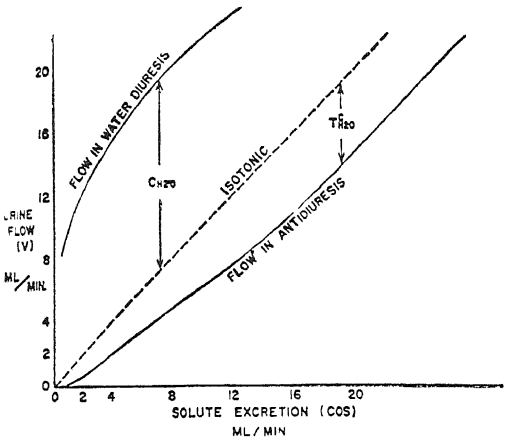
$$C_{os} = U_{os}V/P_{os}$$

When the urine is dilute the flow exceeds the osmolal clearance by an amount which Smith has designated the *free water clearance* (C_{H_2O}) This is the volume of water which must be removed from one minute's flow of urine to return the urine to isotonicity Thus

$$V - (U_{os}V/P_{os}) = C_{H_2O}, \qquad C_{H_2O} = V - C_{os}$$

From the mechanism described above it may be seen that, provided no water has escaped from distal tubule or collecting duct, C_{H_2O} is the volume of water which originally contained the salt reabsorbed in ascending limb, distal tubule and collecting duct The capacity of these segments to reabsorb sodium salts, therefore, sets the limit on the amount of water which can be excreted above that containing the urinary solutes in isotonic solution It is this limited capacity to excrete water freed of solute which accounts for the increasing concentration of dilute urine as solute excretion increases, the isotonic portion of the urine flow (C_{os}) constituting an ever larger fraction of the total flow (V) Actually C_{H_2O} may be less than maximal *1)* because insufficient sodium salts are delivered to the distal portions of the tubule to saturate the reabsorptive capacity or *2)* because some of the water "freed" by salt removal diffuses out of the distal tubule Even in the total absence of ADH some permeability to water persists, and this permeability may be increased by amounts of ADH that are too small to yield a maximal effect and thus a concentrated urine

When the urine is concentrated the flow is



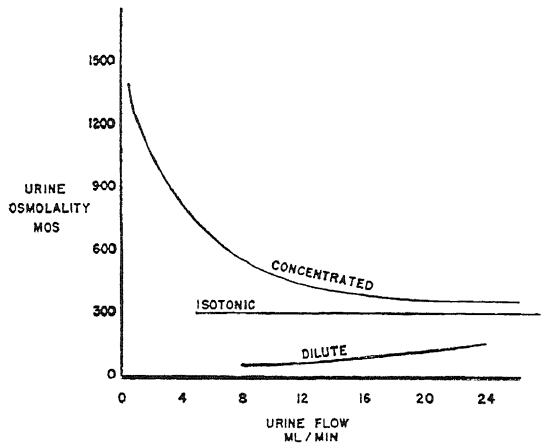


Fig 5 9*a* (*left*) Relationship between urine flow and solute excretion in water diuresis and during the formation of concentrated urine

Fig 5 9*b* (*right*) Relationship between urine flow and urine osmolality when urine flow is modified by changing solute excretion See text for explanation

smaller than the osmolal clearance, that is, water must be *added* to the urine to render it isotonic Thus C_{H_2O} assumes a negative value, this has been designated as $T^c_{H_2O}$ by Smith (1952) The driving force for the removal of water from urine flowing through the collecting ducts is provided by the salt transported out of the ascending limb of the loop of Henle Therefore the amount of water thus removed ($T^c_{H_2O}$) can not be greater than the osmotic equivalent of the salt deposited in the medulla by the loop [10] Since the value of $T^c_{H_2O}$ is limited, as solute excretion increases the urine concentration falls, again because the isotonic portion of the urine (C_{os}) constitutes an ever greater fraction of the total (V) ($V = C_{os} - T^c_{H_2O}$) The flows and concentrations of dilute and concentrated urines as solute excretion is varied are illustrated in Figures 5 9A and 5 9B

In considering the effects of antidiuretic hormone and the change from dilute to concentrated urine, it is important to emphasize that by far the most important effect is that in the cortical segments of the nephron system which involves the change from dilute to isotonic urine The relative quantitative importance of the dilution and concentration processes can be illustrated in a person excreting enough solute to require the excretion of 2 liters of isotonic urine in 24 hr If he were to produce maximally dilute urine for this period, he would excrete approximately 20 liters of urine Therefore, if ADH were to effect a rise in urine concentration only to isotonicity with plasma, some 18 liters of water would be conserved If, on the other hand, the same solute were to be excreted in a maximally concentrated urine, the volume would be reduced to about 500 ml, a further saving of about 1 5 liters Clearly, the water saved in not putting out a dilute urine is quantitatively much more important than that conserved in forming a concentrated urine

[10] This is equivalent to saying that the production of a concentrated fluid (urine) in the medulla must be balanced by the production of a dilute one (the fluid emerging from the ascending limb of the loop)

REABSORPTION AND EXCRETION OF SODIUM AND CHLORIDE

The regulation of the osmotic pressure of body fluids tends to adjust the amount of water in the body to the amount of solute in the body fluids Thus, so long as osmotic pressure regulation remains normal, the volume of the body fluids will depend on the amount of solute in the body Since sodium salts constitute some 90% or more of the solute of extracellular fluid, the volume of the extracellular fluid is largely dependent on the regulation of renal sodium excretion

We have already considered, in relation to the regulation of water excretion, how sodium reabsorption is distributed in the renal tubule It has been pointed out that a major fraction of the filtered solute and water are reabsorbed in the proximal tubule Since most of this solute

is composed of sodium salts and since the fluid remains isotonic in the process of this reabsorption, there is ordinarily little reduction of sodium concentration in the proximal tubule However, the reabsorption involves active sodium transport since *1*) the reabsorption occurs despite a small difference of electrical potential (3 to 4 mv , lumen negative to interstitial fluid) opposing the movement of sodium, and *2*) if some nonreabsorbable solute (e g , mannitol) is present, the sodium concentration does fall well below that of the peritubular fluid as sodium is reabsorbed (However, because of a relatively high permeability the gradient of sodium concentration that can be established is limited because of diffusion from peritubular fluid) Since chloride reabsorption is favored by the small electrical potential, and usually by a higher concentration of chloride in the lumen than in the interstitial fluid, there is no reason to believe any specific transport mechanism for chloride reabsorption is present

The reabsorption of sodium in the more distal segments of the nephron does result in the establishment of gradients of sodium concentration which may, indeed, be very steep since sodium may be virtually absent from the urine The difference between proximal and distal tubules probably resides not so much in the character of the transport mechanism, as in the passive permeability of the tubule epithelium, there is less tendency, in the distal area, for the sodium to diffuse back into the lumen and dissipate the gradient However, in the distal convoluted tubule the removal of chloride may be the consequence of an active transport process since at least one study has been reported as showing that the concentration may fall to levels too low to be accounted for by the electrical force provided by the sodium transport

In addition, in the distal tubule, not all of the reabsorption of sodium is accompanied by reabsorption of anion Some of the sodium reabsorbed in the distal segment is replaced by potassium, and some is replaced by hydrogen and ammonium ions These aspects of the sodium reabsorptive mechanism will be considered in relation to potassium excretion and the renal regulation of acid-base balance in later sections of this chapter.

The mechanism by which sodium is transferred from lumen to peritubular fluid is poorly understood It is probably very closely related to that mechanism (also poorly understood) common to virtually every cell of the body, by which sodium is extruded from and potassium maintained at a high concentration in the intracellular fluid No entirely satisfactory and generally applicable model is available, but it is believed that an ATPase requiring both NA and K for its activation may play an important role

REGULATION OF SODIUM EXCRETION

A number of circumstances are known to influence the rate of sodium excretion The way in which these factors are integrated to determine and regulate the rate of sodium excretion is not so well known To say that the rate of sodium excretion is determined by the balance between filtration of sodium at the glomeruli and reabsorption of sodium by the renal tubules is merely to state the obvious What is required is a knowledge of how these two processes are interrelated and modified to maintain a normal body sodium content Although we can indicate qualitatively some of the determinants of sodium excretion, they can not be put together as, for instance, in the case of glucose, to predict the rate of salt excretion

Glomerular Filtration. Under average normal conditions only a very small fraction of the sodium contained in the glomerular filtrate is excreted in the urine—a person on an average salt intake puts out only about one-half of one percent of the amount filtered. It is clear, therefore, that an increase of only 1% in the rate of glomerular filtration[11] supplies enough additional sodium in the filtrate to double or triple the sodium excretion were none of the extra sodium reabsorbed by the tubules It is also clear, however, that when the glomerular filtration rate increases, the increment of filtered sodium is never quantitatively excreted In fact, unlike the situation with respect to the reabsorption of such substances as glucose and phosphate, changes in the rate of glomerular filtration are attended by parallel changes in the rate of sodium reabsorption

[11] Changes in the amount of sodium filtered are generally due largely to changes in the rate of glomerular filtration since the plasma sodium concentration is normally maintained within a very narrow range, reflecting the constancy of osmotic pressure to which sodium salts contribute such a major share

When glomerular filtration increases, sodium reabsorption increases, when filtration decreases, reabsorption likewise decreases When changes in glomerular filtration occur the reabsorption of salt in the proximal convoluted tubule changes very nearly in exact proportion However, because similar changes do not take place in reabsorption in the remainder of the nephron, changes in the filtration of sodium are not accompanied by proportional changes in excretion As a matter of fact, acute changes in the filtration of sodium produce relatively very large changes in sodium excretion However, if the filtration rate changes are maintained for a protracted period compensatory changes in reabsorption occur so as to restore sodium balance

These circumstances are very clearly demonstrated by the experiments of Mueller and his associates (1951) In dogs operated upon to make it possible to collect the urine from each kidney separately, these investigators placed constricting clamps on one renal artery It was found that if the artery was constricted so as to produce a barely detectable decrease in glomerular filtration rate, there was a marked drop in the amount of sodium excreted by the affected kidney Such a kidney might excrete only 10 to 20% of the daily sodium intake, the remainder being excreted by the normal control kidney so as to maintain sodium balance This situation might persist indefinitely If, however, the normal kidney was then removed there was a prompt increase in the salt excreted by the kidney with the constricted artery When the full load of maintaining sodium balance was placed on the experimental kidney, compensatory adjustments of reabsorption overcame the effects of reduced glomerular filtration

Adrenocortical Steroids. The best known of the influences that modify reabsorption of sodium by the tubules is that exerted by the steroid hormones of the adrenal cortex The natural hormone with the greatest effect on sodium reabsorption is aldosterone, but similar effects are produced by desoxycorticosterone, cortisol, and others, as well as a number of synthetic steroids All of the active steroids exert qualitatively identical effects on sodium transport by the renal tubules, increasing reabsorption in all its forms—that is, sodium reabsorption accompanied by chloride reabsorption as well as sodium exchange for potassium and hydrogen ions However, the various steroids may produce differing effects on sodium *excretion* not only because they differ in potency with respect to their capacity to modify sodium transport, but probably in part because glucocorticoid activity is associated with the specific property of increasing the rate of glomerular filtration

The control of salt excretion by the adrenal cortex is mediated largely by the secretion of aldosterone The secretion of aldosterone is stimulated by angiotensin, a small peptide hormone formed in the blood from a circulating precursor globulin by the action of the enzyme renin Renin is secreted by the juxtaglomerular apparatus of the kidney in response to a stimulus which has not been clearly established, but which appears to be related in some way to the effective circulating blood volume Contraction of blood volume leads to increased secretion of renin which in turn yields increased formation of angiotensin and, in turn, increased secretion of aldosterone, increased reabsorption of sodium (and chloride) and secondarily, by an increase in the retention of water, to expansion of volume

The action of aldosterone is to increase the formation of a protein which plays some as yet unidentified role in the sodium transport process

It should be clearly understood that the transport process proceeds very well in the absence of the corticosteroids, the effect of the hormones is, in a sense, the fine adjustment of a process which is largely autonomous

When adrenal salt-retention hormones are administered to normal individuals there is a brief period of salt retention but the decreased salt excretion is relatively short-lived After retaining a modest amount of salt, the individual comes back into balance This does not indicate the development of resistance to the effect of steroid on sodium transport but compensation for this effect by other factors

Other Controls. In recent years it has become apparent that factors other than glomerular filtration rate and adrenocortical steroids play a dominant role in the regulation of sodium excretion Although these factors are still poorly understood a considerable amount of experimental data that relate to their operation has been amassed If an animal is given massive doses of aldosterone or other salt retention hormones and then infused with iso-

tonic Ringer's solution, a sharp increase in sodium excretion occurs even if the rate of glomerular filtration is markedly reduced by clamping the renal artery or the aorta, the increased salt excretion occurs in denervated kidneys and when the plasma oncotic pressure is maintained by inclusion of albumin in the infused solution (De Wardener et al , 1961) A part of the effect has been shown to be a marked reduction of reabsorption in the proximal convoluted tubule, but additional effects on reabsorption in more distal segments are required for the full expression of the increased sodium excretion

Numerous attempts have been made to identify, in the blood of saline infused animals, humoral depressors of sodium transport by the tubules No conclusive evidence for the existence of such humoral agents has yet been adduced although factors that modify transport in several assay systems have been reported to occur in the plasma of saline infused animals and to be absent in that of hydropenic animals While a regulatory role for some of these agents is a possibility, much work remains to be done before this can be considered established

At the same time it has been shown that the rate of reabsorption by the tubules can be modified by changes in the oncotic and hydrostatic pressures in the peritubular capillaries, presumably through some poorly understood effect of these factors on the rate of transport by the tubules (Martino and Earley, 1967) The regulatory role of such factors also remains to be clarified

Osmotic Diuresis. The excretion of sodium is increased by any circumstance which leads to an increase in the excretion of solute (Wesson and Anslow, 1948) Perhaps the most common such event is elevation of the blood glucose to levels at which the reabsorptive capacity of the tubules is saturated and glucose is excreted in the urine The same effect may be produced by the administration of poorly reabsorbed substances such as urea or mannitol This phenomenon is known as *osmotic diuresis* Sodium excretion is increased because the presence of the other solute depresses sodium reabsorption in the proximal tubule As sodium is reabsorbed, the other solute, which remains behind in the tubule, forms a larger and larger fraction of the total solute Since the fluid retains the same total solute concentration as the blood, the sodium concentration must fall, the decrease in concentration impairs reabsorption

Excretion of Poorly Reabsorbed Anions or Anions Subject to Tubular Excretion When large amounts of some anion are excreted in the urine, there must be an equivalent amount of cation excreted To some extent sodium may be substituted for by potassium, ammonium, or hydrogen ions, but the capacity to do this is limited Thus, under conditions where large amounts of sulfate, phosphate, *p*-amino hippurate (PAH), etc , are excreted, sodium excretion is increased

Specific Inhibition of Sodium Transport. The commonly used diuretics,—organic mercurials, thiazide diuretics, furosemide, ethacrynic acid, carbonic anhydrase inhibitors, etc , all owe their efficacy to the capacity to depress sodium transport by the renal tubules

EXCRETION OF POTASSIUM

Potassium holds a unique place among the normal constituents of the blood and urine in being subject to both reabsorption and secretion by the tubules The average rate at which potassium is excreted in the urine is only about 15% of the rate at which it is filtered at the glomeruli, on the other hand, if large amounts of potassium salts are administered, or if the secretion of hydrogen ion is inhibited, the rate of excretion of potassium may rise to levels equivalent to twice the rate at which it is filtered There is considerable evidence to support the view that, even when the amount excreted is less than that filtered, the potassium in the urine is derived largely from tubular secretion (Berliner, 1961)

The bulk of potassium reabsorption occurs in the proximal tubule The process in this segment if probably passive since the potassium concentration in the convoluted portion does not differ appreciably from that in plasma Reabsorption continues beyond the convoluted portion of the proximal tubule since only a small fraction of the filtered potassium remains when the fluid enters the distal convoluted tubule

The potassium secretory process begins in the distal convoluted tubules and continues into the collecting system The process is, in effect, an exchange of potassium ions for sodium ions, the reabsorption of sodium setting up an electrical gradient with the lumen of the

tubule strongly negative This electrical gradient causes potassium ions to diffuse into the lumen through the distal tubule lining which has a high permeability to potassium Thus sodium reabsorption provides the driving force for potassium secretion This driving force is opposed by active transport of potassium out of the lumen, the actual amount of potassium at any point along the distal system being determined by the balance of these two processes When the supply of sodium is exhausted before the end of the nephron, the driving force provided by sodium reabsorption falls off and potassium is reabsorbed sometimes to very low levels

The rate of potassium excretion is determined by several factors *1)* The concentration of potassium in cells, particularly the renal tubule cells The concentration of potassium in the plasma has little influence on potassium excretion except insofar as cell potassium concentration tends to vary with plasma concentration Such an arrangement is highly reasonable since potassium is primarily an intracellular ion, and it is, therefore, the cell potassium to which excretion should be related *2)* The amount of sodium remaining as the end of the distal tubule system is reached (Davidson et al , 1958) Thus, paradoxically, since sodium is reabsorbed as potassium is secreted, sodium and potassium excretion frequently vary in parallel rather than inversely *3)* The acid-base status In general, alkalosis increases the secretion of potassium and acidosis depresses it The effect, in all probability, is upon the initial step in potassium secretion, the uptake of potassium into the cells at their contraluminal cells (Malnic et al , 1971) The uptake of potassium into cells generally is enhanced by alkalosis and decreased by acidosis *4)* Tolerance to potassium For reasons that are not entirely clear potassium excretion is facilitated by repeated administration of potassium (Thatcher and Radike, 1947)

RENAL REGULATION OF ACID-BASE BALANCE

Under normal conditions the diet provides a number of sources of potential acid The organic phosphorus of proteins and phospholipids is converted to phosphoric acid, the sulfur of the sulfur-containing amino acids is converted to sulfuric acid When ammonium chloride is ingested, the ammonium ion is converted to urea and hydrogen ion leaving hydrochloric acid as the end product If calcium chloride is ingested, the calcium may be excreted in the stool as the insoluble salts of organic acids (such as fatty acids), and hydrochloric acid is absorbed Under abnormal conditons (starvation, diabetic ketosis) large amounts of organic acid may be produced, well beyond the capacity of the body to oxidize this acid to carbon dioxide

In the usual dietary circumstances the dietary sources of acid predominate over the sources of alkali, but the latter may be in excess with certain intakes Fruits and vegetables abound in the salts of organic acids The organic anions are oxidized to CO_2 leaving the bicarbonates of sodium and potassium In addition the intake may include sodium or potassium bicarbonate as such

Acid or alkali entering the body fluids react with the buffers of these fluids and the pH change which occurs is thereby minimized The most abundant buffer system of extracellular fluid is that made up of bicarbonate and carbonic acid and the bulk of the extracellular buffering is therefore involved with this system Because it must provide the major part of the alkali to neutralize any acid gaining access to the extracellular fluid, the bicarbonate concentration has been designated the *alkali reserve* and can be considered a measure of the degree of acidosis or alkalosis (even though, by respiratory compensation, changes of bicarbonate concentration may be accompanied by only minimal changes in pH)

Acid added to extracellular fluid reacts chiefly with bicarbonate to yield carbonic acid and the salt of the acid The carbonic acid in turn yields CO_2 which is excreted in the lungs

$$H^+A^- + Na^+HCO_3^- \rightleftharpoons Na^+A^- + H_2CO_3 \Downarrow H_2O + CO_2$$

The result is the replacement, in the extracellular fluid, of one equivalent of bicarbonate with one equivalent of the anion of the acid The anion (A^-) will eventually be excreted in the urine If it is excreted as the sodium salt, the form in which it exists in plasma, the overall result will have been the disappearance from the extracellular fluid of 1 mole of sodium bicarbonate for each mole of HA added

The buffer capacity of the extracellular fluid can be restored only by regaining a mole of

sodium bicarbonate This might be done if instead of excreting the sodium salt (NaA) in the urine the kidney were able to dispose of the free acid HA thus regenerating sodium bicarbonate This would amount to a reversal of the reactions above

$$H_2O + CO_2 \rightarrow H_2CO_3 + Na^+A^- \rightarrow$$
$$Na^+HCO_3^- + HA \rightarrow (\text{urine})$$

This, in fact, the kidney is able to do by virtue of its capacity to produce urine considerably more acid (down almost to pH 4) than the blood (pH 7 4) Thus to the extent that any acid is un-ionized at the pH of the urine, that acid may be disposed of directly, with regeneration and recovery of the bicarbonate by which it was temporarily buffered in the extracellular fluid

The extent to which an acid is un-ionized at a given pH can be deduced from the Henderson-Hasselbalch equation

$$pH = pK + \log [A^-]/[HA]$$

This may be rearranged in the form pK – pH = log [HA] – log [A^-] to indicate more clearly that when the pH is lower than the pK of the acid concerned there will be more un-ionized acid than salt Thus, any acid which has a pK higher, *or not much lower* than the pH of the urine,[12] can be excreted to an appreciable extent in un-ionized form Acids with pK's high enough to appear in the urine in undissociated form to an appreciable extent are monobasic phosphate ($H_2PO_4^- \rightleftharpoons HPO_4^= + H^+$), pK 6 80, β-hydroxybutyric acid, pK 4 7 and creatinine ($CrH^+ \rightleftharpoons Cr + H^+$), pK 5 0

For most acids which enter the extracellular fluid, excretion as the undissociated acid is impossible Hydrochloric and sulfuric acids are completely dissociated at all pH's At a urine pH as low as 4 0, the concentration of hydrochloric or sulfuric acid would be only 10^{-4} N or 0 1 mEq /liter Even *weak* acids having pK's much below 5 must be excreted almost entirely as salts In the excretion of such acids, then, some other mechanism is required if bicarbonate is to be regained to replace that originally lost in neutralizing them This is accomplished by the generation of ammonium ion from neutral precursors and the substitution of ammonium ion for sodium ion in the urine The acids are excreted as ammonium salts, the sodium is returned to the extracellular fluid as sodium bicarbonate The overall reaction may be considered to be the following

$RNH_2 + HOH \rightarrow NH_3 + ROH$ (which may then be oxidized to CO_2 and H_2O)

$$NH_3 + Na^+A^- + H_2CO_3 \rightarrow$$
$$NH_4^+A^- \text{ (to urine)} + Na^+HCO_3^-$$

Neither of these reactions nor those indicated above should be interpreted as indicating the actual processes which take place in the kidney They represent the net results of a series of reactions

Thus acids are disposed of without loss of alkali reserve *1)* by excretion of weak acids in their un-ionized form and *2)* by the substitution of ammonium ion for sodium ion in the excretion of their salts The disposal of excess alkali is more direct Alkali appears in the extracellular fluid as bicarbonate ($B^+OH^- + CO_2 \rightleftharpoons B^+HCO_3^-$) and is excreted in the urine as that bicarbonate which escapes reabsorption in the tubules This background concerning the nature of the overall processes will serve as an introduction to the renal mechanisms that determine the excretion of acid or alkali

Three processes are important in regulating the amount of acid or alkali excreted *1)* the filtration of bicarbonate at the glomeruli, *2)* the transport mechanism by which bicarbonate is reabsorbed and the urine rendered acid, and *3)* the synthesis of ammonia by the cells of the renal tubule

The glomerular filtrate contains bicarbonate at a concentration essentially identical to that in plasma [13] Micropuncture studies have shown that there is some variation in the pH and bicarbonate concentration as the fluid flows through the proximal convoluted tubule Under normal conditions in the dog, the bicarbonate concentration, and hence the pH, is close to that in plasma When plasma bicarbonate is reduced by acidosis, the bicarbonate concentration falls in the proximal tubule to

[12] Although more acid urine *can* be formed, the urine pH is not often below 5 0 and is more commonly around 6

[13] Because the plasma proteins are nonfilterable anions, a Donnan equilibrium is set up across the glomerular membrane, and all monovalent filterable anions, including bicarbonate, are about 5 per cent more concentrated in glomerular filtrate than in plasma water All monovalent filterable cations, including H^+, are about 5 per cent less concentrated

levels one-half to one-quarter that in plasma Since the tubule has a very high permeability to CO_2, the carbon dioxide tension and, hence, the carbonic acid concentration in the lumen are very close to those of plasma Thus the denominator in the Henderson-Hasselbalch equation,

$$pH = 6.1 + \log([HCO_3^-]/[H_2CO_3]),$$

remains constant and pH varies directly with the log $[HCO_3^-]$ It should be noted that even though the bicarbonate concentration may change but little in the proximal tubule, most of the filtered bicarbonate is usually reabsorbed in this segment as a consequence of the reabsorption of a large fraction of the filtered volume

In the distal tubule the urine is further acidified, and the process continues into the collecting ducts The major reduction in pH occurs in the distal parts of the nephron (Montgomery and Pierce, 1937), but this does not necessarily indicate that the reabsorption of base occurs predominantly in the latter segments As bicarbonate is removed from the tubule lumen, changes in the pH of the remaining fluid will depend upon the changes in volume of the remaining fluid If bicarbonate removal proceeds more rapidly than volume decreases, the bicarbonate concentration falls and the pH likewise falls since the concentration of carbonic acid remains constant (see above) If the volume of the remaining tubular fluid diminishes more rapidly than does the amount of bicarbonate remaining, the pH of the fluid will rise Thus, the degree to which the urine is acidified depends not only upon the rate of bicarbonate removal, but on the relationship of the removal rate to the rate of removal of water Consequently, the production of urine more alkaline than plasma requires only that a greater fraction of the filtered water than of the filtered bicarbonate be reabsorbed

MECHANISM OF URINE ACIDIFICATION

Secretion of Hydrogen Ions Largely as the result of the studies of Pitts and his associates, (1945, 1946, 1948, Sartorius et al , 1949, Pitts, 1948), the process by which bicarbonate removal is effected is generally accepted as being one in which hydrogen ion is secreted by the cells of the renal tubule in exchange for sodium ions The mechanism may be conveniently depicted as follows (Fig 5 10) In the renal tubule cell, as a result of energy-requiring processes, hydrogen ion and hydroxyl ion are generated from water in spatially separated loci The hydrogen ion enters the tubule lumen in exchange for a sodium ion which is taken into the cell Meanwhile the hydroxyl ion is buffered within the cell by reacting with carbonic acid to yield bicarbonate ion and water The bicarbonate and sodium ions are extruded into the extracellular fluid The carbonic acid is regenerated by the hydration of carbon dioxide, the latter may be derived from either the metabolic processes within the cell or from the carbon dioxide of the blood and extracellular fluid

The regeneration of the carbonic acid proceeds rapidly only in the presence of the enzyme *carbonic anhydrase* and this step may become rate-limiting in the overall process When the production of carbonic acid is impaired by inhibition of the enzyme, the concentration of carbonic acid in the cell falls and with it the hydrogen ion concentration (Hober, 1942) Since the rate at which hydrogen ion is extruded in exchange for sodium ion is related to the hydrogen ion concentration in the cell, the rate of secretion of hydrogen ion is diminished Thus *inhibition of carbonic anhydrase diminishes the secretion of hydrogen ion* and all of the processes dependent upon the secretion of hydrogen ion

It should be noted that in the above process, for each hydrogen ion extruded, one equivalent of sodium as the bicarbonate is added to the extracellular fluid and this result is independent of the fate of the hydrogen ion in the

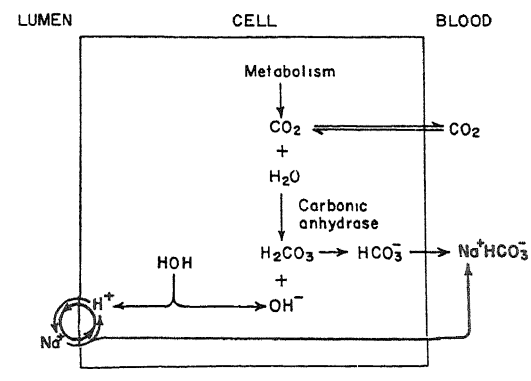


Fig 5 10 Hypothetical cellular processes in the secretion of hydrogen ion in exchange for sodium

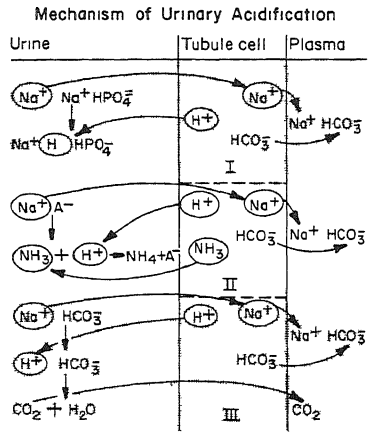


Fig 5 11 Fate of hydrogen ion secreted into tubule lumen in exchange for sodium *a)* Formation of titratable acid *b)* formation of ammonium ion *c)* reabsorption of filtered bicarbonate Modified from figure of Orloff 1956

tubule lumen (Fig 5 11) The fate of the hydrogen ion that enters the lumen in exchange for a sodium ion depends upon the composition of the fluid that it enters

Acid added to any buffered solution is distributed among the buffers of that solution in a predictable way depending upon the relative concentrations of the buffers and their pK values Any change in hydrogen ion concentration must affect all the buffers present since all must satisfy their ionization constants The amount of hydrogen ion distributed to each buffer will be directly proportional to the concentration of the buffer and inversely related to the difference between the pH of the solution and the pK of the buffer

We may consider as an example the secretion of hydrogen ion into the proximal renal tubule The buffers contained in the glomerular filtrate may be considered to be bicarbonate (pK′ 6 1) at a concentration of about 27 mmoles/liter and phosphate (pK 6 8) at a concentration of about 1 5 mmoles/liter Because it is present in much higher concentration, bicarbonate will take up most of the hydrogen ion secreted even though the pH of the fluid (7 4) is closer to the pK of phosphate Thus, in the proximal tubule most of the added hydrogen ion reacts with bicarbonate

$$H^+ + HCO_3^- \rightleftarrows H_2CO_3 \rightleftarrows H_2O + CO_2$$

The result of hydrogen ion addition, then, is the conversion of bicarbonate ion to CO_2 and water, and the CO_2 diffuses out of the tubule lumen to extracellular fluid Thus the exchange of a mole of sodium ion for a mole of hydrogen ion has resulted in the removal of a mole of sodium bicarbonate from the glomerular filtrate even though no bicarbonate ion has crossed the membrane separating lumen from tubule cell Instead, the bicarbonate ion that enters the extracellular fluid is generated independently in the tubule cell

In the distal tubule the situation may be quite different Suppose that the proximal tubule's capacity for phosphate reabsorption has been saturated so that half of the filtered phosphate has escaped reabsorption and reaches the distal tubule At the same time the volume of the fluid flowing through the tubule has been greatly diminished by the processes of reabsorption, so that we may assume that at some point in the distal tubule only 5% of the glomerular filtrate remains The phosphate concentration in the fluid will then have reached a value ten times that in the glomerular filtrate or 15 mmoles/liter Now the hydrogen ion secreted into the fluid will to a very considerable extent react with dibasic phosphate, converting it to monobasic phosphate

$$H^+ + HPO_4^- \rightleftarrows H_2PO_4^-$$

As the pH falls, the concentration of bicarbonate diminishing rapidly, a greater proportion of the hydrogen ion will be taken up by other buffers At pH 6 1, the bicarbonate concentrations will be only about 1 4 mEq /liter, at 5 1 about 0 14 mEq /liter

It is important to note that as hydrogen ion reacts with bicarbonate the resulting undissociated acid (H_2CO_3) disappears from the urine (as CO_2) The hydrogen ion that reacts with other weak acids, however, remains and can be detected in the final urine in the form of *titratable acid* The amount of the latter can be determined as the amount of alkali required to titrate the urine back to the pH of plasma, and is equal to the amount of hydrogen ion expended and sodium bicarbonate regenerated in the conversion of the buffer salt (as it existed in plasma) to the buffer acid as it appears in the urine The titratable acid, therefore, is a very poor measure of the amount of hydrogen ion secreted by the renal tubules since most of the acid produced by this secretion is carbonic acid which is dissipated

However, from the standpoint of the regulation of the acid-base balance of the body, the

situation is reversed That part of the secretion of hydrogen ion which is expended in reabsorbing bicarbonate from the glomerular filtrate serves an important conservative role in preventing loss of bicarbonate from the body, but its overall effect is only to return to the extracellular fluid that which it previously had lost for a time into the glomerular filtrate The excretion of titratable acid, however, represents the addition of extra sodium bicarbonate to the extracellular fluid, replacing the neutral salt (Na^+A^-) which it had previously contained This, then, is an addition to the alkali reserve and as such can be used to balance off the acid products of ingestion and metabolism

The excretion of titratable acid is a process in which there is a net gain of alkali reserve, but it is a process of limited potentiality, limited by the capacity of the kidney to reduce the pH of the urine and by the amount of weak acid of suitable pK excreted in the urine This process alone would be unable to maintain the alkali reserve against the continued addition to extracellular fluid of even most of those acids that can be excreted to a considerable extent in un-ionized form Thus, β-hydroxybutyric acid can be excreted about half as the free acid, half as the salt The excretion of titratable acid can therefore diminish the attrition of body buffer by about one-half, but the continued production of this acid would eventually exhaust the extracellular bicarbonate supply Of course, the production of titratable acid contributes nothing to balancing off the effects of the ingestion or production of strong mineral acids The latter effects are overcome by the production and excretion of ammonium ion in place of so-called "fixed cation" (largely sodium and potassium)

Excretion of Ammonia. The source of urinary ammonia was the subject of debate for many years That the ammonia is produced in the kidney was established by the studies of Nash and Benedict (1921) who showed that more ammonia left the kidney in renal venous blood than entered in the arterial blood The experiments of Van Slyke and his associates (1943) demonstrated that a major fraction of the urinary ammonia was derived from the amide nitrogen of glutamine

$$\text{Glutamine} \rightarrow NH_3 + \text{glutamic acid}$$

The more recent work of Pitts (1964) has shown that much of the remainder is derived from the amino group of glutamine Ammonia can also be produced by the oxidative deamination of a number of amino acids (glycine, alanine, leucine, aspartic acid) but under ordinary conditions the urinary ammonia appears to be derived from glutamine and asparagine with smaller contributions from glycine, citrulline, tryptophan, and proline

Ammonia, produced in the cells of the renal tubules, enters and accumulates in the renal tubules by simple diffusion The nature of the process by which this accumulation takes place may be understood from a consideration of the nature of cell membranes and the solubility characteristics of weak electrolytes (Orloff and Berliner, 1956)

Most cell membranes that have been studied have the characteristics of lipid layers permeated by aqueous pores which occupy only a very small part of the total surface area Substances that are highly soluble in lipid can penetrate the entire surface of the cell, while those that are highly water-soluble and lipid-insoluble can penetrate only through the aqueous pores Lipid soluble materials, therefore, diffuse in and out of cells very much more rapidly than water-soluble substances Among the latter are most ions, ionized compounds are, in general, highly water-soluble and poorly soluble in lipid On the other hand, many un-ionized substances are highly soluble in lipids (provided they are not highly polar compounds, e g , sugars, urea, short-chain alcohols, etc) In accord with these generalizations, the ammonium ion (NH_4^+) is highly water soluble and would be expected to penetrate cell membranes poorly, on the other hand, free ammonia (NH_3) is highly lipid-soluble and penetrates cells quite readily

Now suppose we have two aqueous solutions of different pH separated by a membrane having the characteristics of the lipid layer discussed above To such a system ammonia is added The ammonia, being highly lipid-soluble, will penetrate the membrane readily and will reach equilibrium when the concentration of the free ammonia (NH_3) is the same on both sides of the membrane However, in each solution ammonia will react with hydrogen ion to yield ammonium ion

$$NH_3 + H^+ \rightleftharpoons NH_4^+$$

This reaction will be continuously at the equi-

librium required by the dissociation equation

$$K_A = [H^+][NH_3]/[NH_4^+]$$

or

$$pH = pK_A + \log [NH_3]/[NH_4^+], \qquad pK_A = 8.9$$

Since $[NH_3]$ is the same on both sides of the membrane, the numerator of last term may be considered a constant and the log $[NH_4^+]$ will vary inversely with the pH Thus, a decrease in pH of one unit will yield a 10-fold increase in ammonium ion concentration Since the free ammonia constitutes a very small part of the total ($NH_3 + NH_4^+$) the total will vary pretty much as the concentration of ammonium ion

This model may be considered to approximate the situation in the kidney where the solution on one side of the membrane represents the contents of the renal tubule cells in which ammonia is produced and the other side of the membrane, the fluid in the lumen We may consider the pH of the intracellular fluid and its ammonia concentration to be maintained relatively constant and ammonia to accumulate in the lumen until the free ammonia concentration is approximately the same on both sides and the ammonium ion concentration in the lumen that required by the urinary pH The concentration of ammonium ion will therefore increase as urine pH decreases

Because several assumptions involved in this model are only approximations, the actual ammonia excretion varies with pH appreciably less than would be predicted (Orloff and Berliner, 1956) A urinary pH change of two to two and a half units is required to yield a 10-fold change in ammonia excretion (The idealized model would indicate a tenfold change with each pH unit) It should be noted that as NH_3 diffuses into the urine, hydrogen ion is titrated and the pH of the urine tends to become elevated, the pH difference would disappear if it were not maintained by the continuous secretion of hydrogen ion

Ammonia appears in the tubule lumen wherever the tubule fluid has a pH low enough to permit the trapping of ammonia as ammonium ion It thus appears in the proximal tubule when the pH falls appreciably below that of plasma (normally in the rat, in metabolic acidosis in the dog) As urine pH approaches its minimum in the distal nephron, each hydrogen ion secreted into the lumen must be neutralized by capture of NH_3 to form an ammonium ion before another hydrogen ion can be secreted

The excretion of ammonia like the excretion of titratable acid represents replenishment of the body's alkali reserve Quantitatively it represents, under most conditions, the more important contribution to the maintenance of the acid-base balance of the body The sum of titratable acid and urinary ammonium ion represent the net gain of sodium bicarbonate for the body fluids as a result of the process of urine formation (In estimating body balance, it is necessary to subtract from this total any bicarbonate excreted in the urine)

A striking and unique feature of ammonia formation and excretion is the phenomenon of *adaptation* If an individual is given an acidifying regimen (such as the administration of ammonium or calcium chloride or the feeding of a strongly acid-ash diet), there is a gradual increase, over a period of three to five days, of the excretion of ammonia in the urine This increase is independent of changes in urinary pH (which occur very much more rapidly) and the amount of ammonia in the urine is eventually greater at each pH than it was before the acidifying stimulus was started The administration of alkali has an effect exactly the reverse of the administration of acid Neither the nature nor the immediate stimulus to this adaptation are too clearly established It undoubtedly involves a modification in the enzymatic mechanism by which ammonia is produced and presumably a rise in the concentration of ammonia in the cells in which it is produced

In the rat, but not in the dog, it can be shown that glutaminase activity is increased (Davies and Yudkin, 1952, Rector and Orloff, 1959) The overall behavior of ammonia excretion in man resembles that in the dog more closely than in the rat For instance, if acidosis is produced by the administration of inhibitors of carbonic anhydrase which cause massive loss of bicarbonate in the urine, ammonia excretion and glutaminase activity increase markedly in the rat, but ammonia excretion remains at a level related only to urine pH in both man and dog There is no evidence of adaptation of the sort produced by the administration of acidifying salts

The stimulus to adaptation is also not clear

It is obviously in some way related to the intake of acid, but it does not appear to depend on the degree of acidosis, ammonia excretion remaining high when, as a result of the adaptation, plasma bicarbonate has been restored essentially to normal

Quantitative Aspects of Urine Acidification As with all processes of urinary excretion, the urinary acidification is dependent upon the balance between glomerular filtration and tubular reabsorption In this case, the important characteristic of the filtrate is its bicarbonate concentration We will postpone consideration of those factors which modify the secretion of hydrogen ion by the renal tubules and consider first the effect of changes in plasma bicarbonate concentration

We must first take note of the fact that for reasons that have not been determined, the capacity of the tubules to reabsorb bicarbonate varies with the rate of glomerular filtration but not with the plasma bicarbonate concentration The concentration of bicarbonate in the glomerular filtrate is therefore a far more important determinant of bicarbonate excretion than the total amount of bicarbonate filtered To take this into account and to present the reabsorption and excretion of bicarbonate as an orderly, predictable phenomenon, it is the practice to divide the amounts filtered, reabsorbed and excreted by the rate of glomerular filtration and to present the data as the amount reabsorbed, etc , per 100 ml or per liter of glomerular filtrate

When the plasma bicarbonate concentration is low, the urine is acid and essentially free of bicarbonate, all of the filtered bicarbonate is reabsorbed As the plasma bicarbonate concentration is increased, the urine remains essentially free of bicarbonate until the concentration approaches approximately 28 mEq /liter Bicarbonate then begins to appear in the urine and as the plasma concentration increases further bicarbonate reabsorption remains virtually constant at 28 mEq /liter of glomerular filtrate, and the excess is excreted in the urine This is illustrated in Figure 5 12 This tends to produce stabilization of the plasma bicarbonate concentration at its normal level of 27 to 28 mEq /liter When the plasma bicarbonate falls below this level, the filtered bicarbonate is completely reabsorbed and conserved If the plasma bicarbonate rises appreciably above 28 mEq /liter, the excess is excreted and the plasma bicarbonate tends to return rapidly to normal

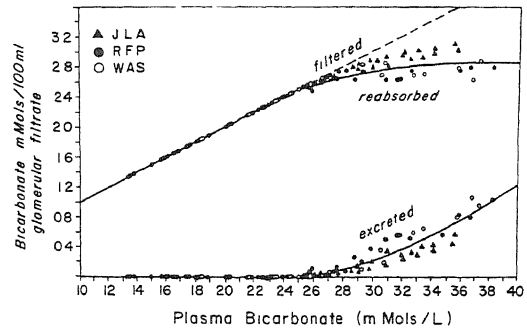


Fig 5 12 Relationship between plasma bicarbonate and reabsorption and excretion of bicarbonate in man (Pitts Ayer and Schiess 1949)

It should be noted that, to the extent that the capacity of the tubule to secrete hydrogen ion is expended in the reabsorption of bicarbonate, no appreciable amount is available to appear as titratable acid or ammonia Thus, when the plasma bicarbonate is close to 28 mEq /liter, essentially all of the filtered bicarbonate will be reabsorbed but there will be little titratable acid and ammonia in the urine (reflecting the fact that the urine pH will not be very low) However, when the plasma bicarbonate drops lower the bicarbonate of the glomerular filtrate may be exhausted before the capacity to secrete hydrogen ion is fully utilized, and a variable part of the remaining capacity will appear as titratable acid (depending on the rate of excretion of buffer acid) or as ammonium ion (depending upon the rate of production of ammonia)

It is also noteworthy that the amount of titratable acid and ammonia in the urine is a poor indication of the capacity of the tubules to secrete hydrogen ion because, if plasma bicarbonate is low, hydrogen ion secretion stops when the urine pH gets too low and before all the capacity for hydrogen-sodium exchange is utilized Thus, the *capacity* to secrete hydrogen ion can be evaluated only when there is bicarbonate left over from the process—that is, when the urine is alkaline

Factors Modifying the Rate of Secretion of Hydrogen Ion. Under ordinary conditions of acid-base balance, with a normal carbon dioxide tension in the blood, a normal cell potassium concentration, etc , the capacity to secrete hydrogen ion in exchange for sodium is

about 28 mEq /liter of glomerular filtrate as indicated in the preceding discussion This tends to maintain the plasma bicarbonate concentration at about 28 mEq /liter The capacity to secrete hydrogen ion may be varied in response to several factors

Carbon Dioxide Tension When the carbon dioxide tension is increased (respiratory acidosis) the plasma bicarbonate concentration increases by titration of the blood buffers, and the pH change attributable to the increased carbonic acid concentration is minimized The elevated plasma bicarbonate concentration persists, and the maintenance implies an increase in bicarbonate reabsorption (hydrogen ion secretion) by the renal tubules Indeed, such an increase in bicarbonate reabsorption is demonstrable immediately upon elevation of the carbon dioxide tension (Relman et al , 1953, Brazeau and Gilman, 1953) If elevation of the carbon dioxide tension is maintained for extended periods there is a further increase in the capacity to secrete hydrogen ion, the nature of this further adaptation is unknown (Sullivan and Dorman, 1955) Decreases in carbon dioxide tension (respiratory alkalosis) have exactly the opposite effects, reducing renal bicarbonate reabsorption (hydrogen ion secretion) and causing stabilization of the plasma bicarbonate concentration at a reduced level

The effects of changes in carbon dioxide tension on hydrogen ion secretion are readily interpreted in terms of the hypothesis that the rate of secretion of hydrogen ion is directly related to the hydrogen ion concentration of the tubule cells The concentration of dissolved CO_2 is directly related to the carbon dioxide tension (P_{CO_2}) and the concentration of carbonic acid (H_2CO_3) is directly proportional to the P_{CO_2} Since cell membranes have a very high permeability to CO_2, all changes in extracellular P_{CO_2} are reflected by similar changes in intracellular P_{CO_2} Thus intracellular pH falls when extracellular P_{CO_2} is elevated and vice versa (On the other hand, changes in extracellular bicarbonate concentration are probably *not* regularly reflected by parallel changes in intracellular bicarbonate and pH)

Potassium Metabolism The secretion of hydrogen ion is depressed when potassium is administered so that the urine frequently becomes alkaline even though the plasma bicarbonate concentration is low Conversely, in the presence of potassium depletion the rate of secretion of hydrogen ion is enhanced so that the plasma bicarbonate concentration becomes stabilized at an elevated level

The effects of potassium on hydrogen ion secretion presumably reflect the fact that the rate of secretion of hydrogen ion is dependent upon the pH of the secreting cells and that when cells lose potassium they become more acid while when they gain potassium they become less acid (Orloff et al , 1953)

Chloride Balance When loss of chloride has been incurred in the course of elevation of the plasma bicarbonate, the excretion of the excess bicarbonate must lead to a contraction of extracellular fluid volume unless chloride is available from the intake to replace the bicarbonate In the absence of such chloride intake, urine acidification continues at a high level

Carbonic Anhydrase The role of carbonic anhydrase and the effect of inhibition of this enzyme have been discussed earlier Under normal conditions the activity of carbonic anhydrase is not rate-limiting in hydrogen ion secretion although it may become so when inhibited There is no evidence that adaptation to acid loads involves an increase in kidney carbonic anhydrase

Adrenal Steroids The secretion of hydrogen ion participates in the enhancement of sodium reabsorption produced by aldosterone and other steroids with similar effects on electrolyte metabolism

PASSIVE REABSORPTION THE EXCRETION OF UREA

Because passive reabsorption is the result of the active reabsorption of other constituents of the glomerular filtrate, it is appropriate to consider this relatively simple mechanism last The reabsorption of urea has long been considered a typical example of this process [14] Certainly as the most abundant constituent of the urine and as one which fits in unique fashion

[14] There is evidence that in ruminants (sheep, camel) there is some sort of active renal conservation of urea when the animal receives a low protein intake Active reabsorption from the collecting duct has also been demonstrated While it has also been suggested that there is a similar mechanism in man, the evidence is not convincing, and for the sake of didactic clarity, any such process will be disregarded

into the formation of highly concentrated urine (Gamble et al, 1934), urea requires separate consideration

Cell membranes generally have a high permeability to urea It is therefore not surprising that as the filtrate is reabsorbed from the renal tubules, tending to concentrate the urea in the remaining fluid, urea tends to diffuse out of the lumen What is more remarkable is the very considerable extent to which the urea concentration can be raised in the urine This appears to be due to two factors *1)* an unusually low permeability to urea of the cells lining some segments of the renal tubules (it is to be remembered that much of the tubule lining membrane also has a uniquely low permeability to water, especially in the absence of antidiuretic hormone), and *2)* a mechanism for maintaining a high urea concentration in the interstitial fluid (of the medulla) surrounding those portions of the nephron that contain fluid with the highest urea concentrations (the collecting ducts), in the latter case the concentration gradient driving the diffusion of urea out of the urine is minimized (Levinsky and Berliner, 1959)

As is characteristic of substances excreted by mechanisms involving passive reabsorption, the fraction of the filtered urea excreted in the urine varies with the urine flow (Moller et al, 1929, Shannon, 1936) The fraction of the filtered urea excreted is given by the ratio of the urea clearance (C_U) to the glomerular filtration rate At urine flows in excess of about 2 ml/min in individuals with normal renal function, there is relatively little further increase in the urea clearance with further elevation of urine flow up to the limits of *water diuresis* At this time the urea clearance averages about 60 to 70% of the inulin clearance, i e, about 30 to 40% of the filtered urea is reabsorbed If *osmotic diuresis* is superimposed, the urine flow can be further increased, and under these conditions the excreted fraction of filtered urea increases further along a curve which may be extrapolated to suggest that no urea would be reabsorbed when no water was reabsorbed (Shannon, 1938) Conversely, as urine flow decreases in antidiuresis the excreted urea diminishes to what may be only 15 or 20% of the filtered These phenomena are illustrated in Figure 5 13 from the data of Shannon obtained in the dog

As was noted earlier, in the discussion of water excretion and reabsorption, the change from the low flow of concentrated urine to the high flow of water diuresis involves primarily modifications in the flow in the distal portions

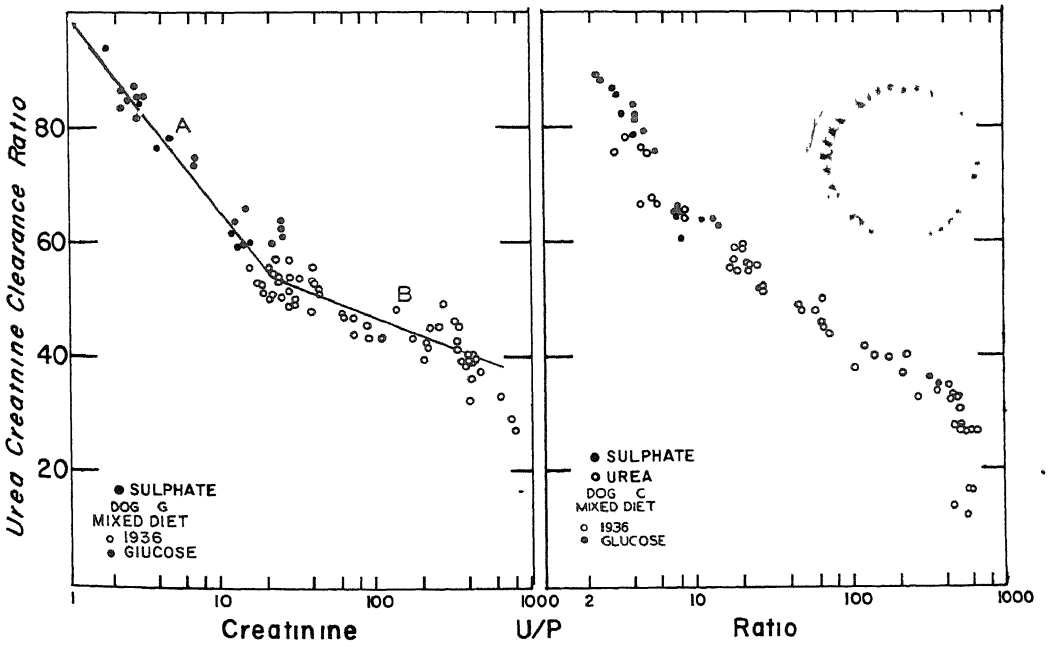


Fig 5 13 The ratio of urea clearance to creatinine clearance (Fraction of the filtered urea excreted in relation to the reabsorption of water as indicated by the U/P ratio of creatinine (Shannon 1938)

of the nephron The reabsorption of water that occurs in the proximal tubule is relatively unaffected Thus, the additional urea that is reabsorbed when water diuresis is interrupted is logically considered to be reabsorbed in the distal tubule and collecting system Furthermore, since in the absence of antidiuretic hormone these portions of the nephron have a low permeability to water, they may be expected to have a low permeability to urea as well It is thus logical to attribute the reabsorption of some 30 to 40% of the filtered urea to losses from the proximal tubule as water is reabsorbed from that segment and to consider the losses from the distal portions of the nephron to be negligible in water diuresis As the fraction of the filtrate reabsorbed in the proximal tubule is diminished in *osmotic* diuresis, the urea is concentrated to a lesser extent and less than the usual 30 to 40% is reabsorbed

In antidiuresis, the reabsorption of water in the distal segments of the nephron system is markedly increased, at the same time, the permeability of the collecting duct to urea is also enhanced Both of these factors lead to an increase in the diffusion of urea out of the urine in the distal portions of the nephron and account for the rapid falling off of the urea clearance as urine flow decreases to low levels

Because the arrangement of the blood supply in the medulla constitutes a countercurrent exchanger, urea escaping from the collecting ducts is not easily carried away in the blood, as it is in the cortex Instead, even a relatively slow leak of urea from the collecting ducts is sufficient to yield a high urea concentration in the peritubular region As a result, the urea concentration in the interstitial space of the papilla is almost as high as in the urine This has several important consequences *1*) The amount of urea lost from the urine is reduced because the concentration gradient is small *2*) The osmotic effect of the urea in the collecting duct urine is counterbalanced largely by urea in the interestitial space, the sodium chloride of the interstitial space, which initiates the urine concentrating process can be utilized to balance the osmotic effect of the non-urea solute, the urea being simply added to both sides of the intervening membrane Consequently, the urine can be made considerably more concentrated when urea is a major part of the excreted solute, and urea, the major product of protein catabolism, and, in fact, by far the most abundant product of metabolism requiring urinary excretion, can thus be excreted largely in water already obligated for the excretion of other urinary solutes *3*) Some of the urea enters the descending limb of the loop of Henle and recirculates through the distal tubule and collecting system In fact, the micropuncture studies of Lassiter, Gottschalk, and Mylle (1961) have shown that, in the antidiuretic rat, the urine flowing through the loops of Henle gains an amount of urea equivalent to about half of the amount filtered so that it would appear that a large part of the urea that escapes the collecting ducts recirculates through the distal tubule and that very little can leave the medulla via the blood in the vasa recta

Because the level of the blood urea and the urea clearance are frequently used for the clinical evaluation of renal function, a brief discussion of these clinical indices is warranted Since the concentration of urea in red blood cell water is the same as in the plasma, the concentration of urea in whole blood is only slightly lower than in plasma, the difference being due to the larger fraction of solids in the red cells Consequently, it has been the usual practice to base the clinical estimation of urea clearance on the measurement of urea in whole blood (although glomerular filtration operates only on plasma) Because of the variation of urea clearance with urine flow some account must be taken of this factor The convention which has been adopted has been to arbitrarily divide urea clearances into two classes When the urine flow exceeds 2 ml /min the clearance of urea is calculated in the conventional fashion except for the substitution of B, the concentration in whole blood, for P, the concentration in plasma The figure (UV)/B obtained is called the *maximal urea clearance* and a value of 75 ml /min is accepted as the average normal It is sometimes the practice to divide the clearance obtained by 75 and express the result as a per cent of normal

When the urine flow is less than 2 ml /min, some correction is necessary to take into account the marked effect of urine flow change on urea excretion It has been found empirically that, in normal individuals, if the clearance is divided by the square root of the urine flow ($\sqrt{V}$) the value obtained is relatively independent of urine flow provided the latter is less than 2 ml /min This figure is referred to

as the *standard urea clearance* and is equal to

$$(UV/B) \times (1/\sqrt{V}) = U\frac{\sqrt{V}}{B}$$

It has an average normal value of 54 and, again, the result obtained is often expressed as a percent of the normal value

When urea clearances are measured, every attempt should be made to assure a urine flow in excess of 2 ml /min since the maximal urea clearance is more clearly related to the value which one wishes to estimate, the rate of glomerular filtration

The blood urea concentration is also, in a rough way, related to the rate of glomerular filtration However, it is subject to modification by factors that do not affect the urea clearance, namely the intake and metabolism of protein and other nitrogen sources For this reason the urea clearance is a much more reliable index of renal function

For the individual in nitrogen balance, almost the entire nitrogen intake is excreted in the form of urea The excretion of urea is in the long term equal to the formation of urea Therefore, if we were to consider the daily excretion of urea as a sort of 24-hr determination of urea clearance we would have

$$C_U = UV/B = \text{rate of nitrogen turnover}/B$$

and

$$B = \text{rate of nitrogen turnover}/C_U$$

Thus, we see that the blood urea is inversely related to the filtration rate to the extent that the urea clearance is related to the filtration rate, but in addition it is also directly proportional to the nitrogen turnover The latter factor is eliminated when the urea clearance is determined

NON-IONIC DIFFUSION

The excretion of weak electrolytes is frequently modified by diffusion dependent upon the dissociation characteristics of the substance concerned The phenomenon is entirely analogous to that involved in the entrance of ammonia into the urine, a process which was discussed in some detail above It was pointed out that un-ionized weak electrolytes frequently penetrate cell membranes very much more readily than their ions As a result when cell membranes separate two aqueous phases of differing pH, the weak electrolyte accumulates in that phase in which the pH requires that the ratio of ionized to un-ionized weak electrolyte is higher The same considerations apply to a number of other weak electrolytes, not only weak bases such as ammonia but weak acids as well Most of the substances the excretion of which has been shown to involve this type of mechanism are foreign to the body, but it is highly probable that a number of normal constituents are also involved although they have not been studied in detail Some of those drugs which have been shown to be excreted by mechanisms involving processes of nonionic diffusion are bases (quinine, atabrine, neutral red, mecamylamine) and acids (salicylates, phenobarbital) As opposed to the situation with respect to ammonia, which is produced in the tubule cells, the two phases of differing pH involved in excretion of these foreign materials are the urine and the peritubular interstitial space

The effect of pH on the excretion of weak acids is exactly the opposite of its effect on that of weak bases This is due to the fact that the ratio of ion to un-ionized fraction increases with pH in the case of weak acids, as indicated by the equation

$$pH = pK_A + \log [A^-]/[HA]$$

while in the case of weak bases the ratio of ionized to un-ionized decreases with rising pH as indicated by the dissociation equation for a base

$$pH = pK_A + \log [RNH_2]/[RNH_3^+]$$

The effect of pH on the excretion of weak bases is appreciable only when the pH of the urine is lower than the pK_A of the base, since it is only when pH is less than pK that the amount of ion is the major fraction of the total For similar reasons pH change has an appreciable effect on acid excretion only when the urine pH is higher than the pK of the acid

The bases and acids noted above as being excreted by mechanisms involving non-ionic diffusion are each secreted into the proximal tubule by active transport processes and escape to a variable extent, depending on urine pH, by non-ionic diffusion in the more distal segments of the nephron

2 / PATHOLOGICAL PHYSIOLOGY OF THE KIDNEY. MICTURITION

Clinical Evaluation of Renal Function

EXAMINATION OF THE URINE

Ordinarily, the first step in the evaluation of the functional status of the kidney, and frequently the source of the first clue to the presence of renal disease, is the clinical pathological examination of the urine This has the unique advantage of extreme simplicity, requiring only a minimum of time and the simplest of equipment Furthermore, the presence of renal disease is usually manifested by some abnormality of the urine However, it is important to note that an apparently normal urine does not exclude the presence of certain diseases of the kidney nor impairment of renal function of even very marked degree

Although the full routine examination of the urine includes several other tests usually intended to detect nonrenal metabolic abnormalities (diabetes, ketosis, etc), those particularly related to disease of the kidney itself are tests for protein and examination of the urinary sediment In addition, determination of the specific gravity and acidity may, under certain conditions, be revealing of disorders of kidney function Certain specific defects of renal function may be detected by analysis of urine for amino acids

TESTS INVOLVING EXAMINATION OF THE BLOOD ONLY

The most commonly used test intended to evaluate the functional state of the kidney is the determination of the blood urea (The results of this determination are commonly reported as the blood urea nitrogen, since nitrogen makes up just under half of the molecular weight of urea, the blood urea concentration is approximately twice the concentration of blood urea nitrogen) The concentration of urea in the blood is directly proportional to the nitrogen turnover and inversely proportional to the renal clearance of urea Its relationship to the nitrogen turnover and to glomerular filtration are fully discussed in the preceding chapter

Since the blood urea concentration is only partially related to the rate of glomerular filtration and to the functional status of the kidney, it is hardly surprising that there is a wide range of normal values for the blood urea concentration and it is likely to be unequivocally elevated only when renal function is reduced by about half or more

Nonprotein Nitrogen Under normal conditions urea contributes about half of the nonprotein nitrogen of plasma (or serum) The remainder is made up of the nitrogen of a heterogeneous group of substances creatine, creatinine, uric acid, amino acids, etc As renal function is depressed the nonprotein nitrogen concentration of plasma rises, largely because of increasing urea concentration, though some of the other components also accumulate The nonprotein nitrogen determination is thus subject to very much the same interpretation and qualifications as the blood urea analysis

Creatinine The plasma creatinine concentration is a somewhat more reliable index of renal function than the blood urea concentration because the rate of formation of creatinine in the body is much more stable than the formation of urea However, some reservation is required because the usual analysis of plasma for creatinine involves a chemical method affected by other constituents of the plasma, this is a more important consideration when the concentration of creatinine is at or near a normal level than when it is substantially increased

TESTS INVOLVING COLLECTION OF URINE

A general comment and caution concerning these tests as a group is warranted These tests (except for the concentration test) involve determination of the rate of excretion of some substance They thus require a *timed collection of urine* and can be no more accurate than the timing and completeness of collection In the case of substances such as urea or creatinine which are normal constituents of the urine, the urine formed before the actual period of the test will contain the test substance and it is, therefore, just as important that the bladder be completely emptied at the start of the test period as that all of the urine be collected at the end Failure to empty the bladder

before the test period will give a result that is misleadingly high just as failure to empty at the conclusion of the test will give an erroneuosly low value Failure to give due care and consideration to this point frequently vitiates the usefulness of what is fundamentally a valuable method for assessing kidney function

Urea Clearance The mechanism of urea excretion has been considered in detail in the preceding chapter, and little need be added here The test is useful for the information it gives concerning the level of glomerular filtration In contradistinction to the simple determination of blood urea, the urea clearance is not dependent upon the rate of formation of urea so that this source of variation is eliminated The urea clearance depends only upon the rate of glomerular filtration and the urine flow When the test is carefully performed it is a very useful index of glomerular filtration rate

Creatinine Clearance The clearance of endogenous creatinine is frequently a valuable index of glomerular filtration As compared with the urea clearance it has the advantage of being unaffected by the rate of urine flow [1] However, for reasons that have not been fully elucidated, the clearance of endogenous creatinine is sometimes considerably greater than the glomerular filtration rate and sometimes lower This is particularly true in the presence of renal disease The creatinine clearance can, therefore, be considered only a rough approximation of the rate of glomerular filtration

Excretion of Phenosulfonphthalein (PSP, Phenol Red) Phenolsulfonphthalein (PSP) is one of that group of organic acids rapidly secreted into the urine by the renal tubules (see preceding chapter) In contrast to some of the other substances in this group, PSP is *not* completely removed from the blood which perfuses the kidney Nevertheless, the rate at which it is excreted by the kidney is determined largely by the rate at which PSP is delivered to the kidney by the flow of blood The excretion of PSP is, therefore, reduced when the renal blood flow is diminished The greatest virtue of the test is its simplicity, PSP is a dye which, in alkaline solution, has an intense red color, and its concentration in the urine is easily estimated

A number of different procedures for the performance of the PSP test have been used, involving various routes of administration of the dye and varying time periods for the collection of the urine Procedures involving intravenous administration of the dye and relatively short periods for the collection of the urine are to be preferred as most sensitive to changes in kidney function Low results can be obtained because of retention of PSP in the bladder as well as because of poor renal function

Concentration Test The concentration test is a measure of the ability of the kidneys to form hypertonic urine Because the test as usually performed involves determination of the urine specific gravity, it can be considered only a rather crude index of this particular function [2] The determination of specific gravity is, however, such a simple procedure that it is unlikely that it will soon be replaced by an estimation of osmolality by the determination of the freezing point depression

In the performance of the test, it is usual to determine the specific gravity of the urine formed after all fluids have been withheld for 10 to 24 hr Normally the specific gravity should be greater than 1 025 although a specific gravity between 1 020 and 1 025 can not be interpreted as a clear indication of abnormality

Clearances of Inulin and *p*-Aminohippurate. Properly performed these measurements yield accurate estimates of glomerular filtration rate and renal blood flow and hence are highly valuable in accurately defining the status of renal function The theoretical basis of the measurements is discussed in the preceding chapter Unfortunately the determination of inulin and *p*-aminohippurate clearances requires the administration of continuous intravenous infusions at constant rate The desired and attainable accuracy can be achieved only with meticulous care in the collection of the urine, usually with bladder catheterization and washout of the bladder at the termination of each collection period Finally, the chemical analysis, particularly that for inulin, is relatively involved For these reasons these tests are rarely used for routine clinical purposes, for most such situations one or more of the approximation procedures may be used

Although the development of isotopically labeled inulin (^{14}C or ^{3}H) and Diodrast (^{131}I) has eliminated many of the problems of chemical analysis in the measurement of filtration rate and blood flow, the other inconveniences remain The precision possible with these methods, properly applied, is rarely needed for clinical purposes and the approximation methods generally suffice Additional approximation methods based upon the disappearance of substances labeled with radioactive markers have been developed The disappearance of a suitable substance (one excreted only by the kidneys and susceptible to la-

[1] However, as in any test involving the collection of urine, the *errors* due to incomplete or inaccurate collection of the urine are minimized if the rate of flow is high

[2] In the formation of hypertonic urine the nature of the urinary solute is more or less a matter of indifference, the process being involved with modifying the urine osmotic pressure to which each solute particle, independent of its nature, contributes equally However, each solute particle does not modify the specific gravity in the same way, one molal solutions of different solutes may vary widely in specific gravity

being with a radioisotope) can be followed by measuring its concentration in serially collected blood samples or, in some instances, even by external counting

MEASUREMENT OF TUBULAR REABSORPTIVE AND SECRETORY CAPACITY

These indices of renal function determined respectively, as the maximum capacity to reabsorb glucose (Tm_G) and to secrete *p*-aminohippurate (Tm_{PAH}) are, like the measurement of inulin and PAH clearances, more or less definitive measurements of renal function (see preceding chapter) Like the latter determinations, the procedures are involved, poorly suited to routine use, and rarely performed for purely clinical purposes

Classification of Renal Diseases

Disturbances of renal function may be separated into two groups—those due to disease of the kidney itself and those due to nonrenal disturbances In the latter group fall abnormalities of renal function due to circulatory disorders and to disturbed hormonal balance with loss or excess of endocrine factors that modify the function of the kidney Among disturbances due to disease of the kidney itself two groups may be recognized *1*) disorders due to specific abnormalities of one or more of the transport mechanisms of the renal tubule, although these are particularly interesting from the point of view of clinicophysiological correlations, they are neither as frequent nor as important as 2) the inflammatory and degenerative diseases which involve the kidney more diffusely, affecting entire nephrons or nephron segments and their functions

In the group of diseases that produce such diffuse involvement of the kidney, a few warrant brief discussion here to provide a background for some of the correlations between disturbances of physiology and renal disease By far the most important of these from the standpoint of frequency of occurrence are glomerulonephritis, arteriolar nephrosclerosis, and pyelonephritis

GLOMERULONEPHRITIS

Glomerulonephritis is a diffuse inflammatory disease involving primarily the glomeruli and secondarily, with loss of glomerular function, degeneration of the conjoined tubules A number of classifications of glomerulonephritis, based largely on pathological distinction, have been proposed and a number of variants are recognized by some workers For present purposes it is sufficient to note that glomerulonephritis can be divided into two forms which may or may not represent different stages of the same disease

Acute glomerulonephritis is characterized by a more or less sudden onset with proteinuria and hematuria, and varying degrees of edema, hypertension and renal insufficiency The onset of the disease can usually be related to an antecedent infection most often with group A, β-hemolytic streptococcus Glomerulonephritis is generally believed to be a disease, involving, in some still obscure fashion, an antigen-antibody reaction which produces damage to the glomerular capillaries Acute glomerulonephritis has a relatively low mortality The great majority of cases recover completely, with no residual evidence of renal disease or of significant functional loss A small proportion do not recover, but show persisting evidence of active renal disease and progressive loss of renal function with an eventually fatal outcome

Chronic glomerulonephritis is a progressive disease characterized by proteinuria, hypertension and diminution of renal function Whether or not it is the progressive form of the same disease recognized as acute glomerulonephritis is disputed In most patients with the chronic form no episode of acute glomerulonephritis is recognized The rate of progress of the disease is extremely variable, some patients have a rapid downhill course with a precipitous decline in renal function Others may, for years, have proteinuria as the only evidence of smoldering disease

THE NEPHROTIC SYNDROME

The nephrotic syndrome is characterized by massive proteinuria leading to a marked reduction of the plasma protein (particularly albumin) concentration, with edema, frequently massive, and generally a marked increase in the concentration of lipids, particularly cholesterol, in the plasma Hematuria and hypertension are usually not prominent A small proportion of instances of the nephrotic syndrome are clearly due to recognizable causes such as *amyloidosis* of the kidney, *lupus erythematosus*, *syphilis*, etc., but in most cases the underlying cause is not recognizable. No unanimity exists as to whether or not the idi-

opathic nephrotic syndrome represents a stage of chronic glomerulonephritis A certain number of patients with the nephrotic syndrome do not recover, but show progressive loss of renal function with the development of hypertension and renal insufficiency in a fashion not distinguishable from chronic glomerulonephritis On the other hand, with antibiotic treatment and prevention of intercurrent infection and with control of the disease with adrenocortical steroids, a considerable proportion of patients with the idiopathic nephrotic syndrome recover completely

Whereas it was formerly believed that in true nephrosis the glomeruli were normal and the disease limited to degenerative changes in the tubule epithelium, this is now recognized as being the exact reverse of the true situation The primary defect is in the capillary basement membrane, permitting the escape of protein into the glomerular filtrate The striking changes in the appearance of the tubule cells on microscopic examination are probably secondary to the high concentration of protein in the glomerular filtrate with protein uptake from the tubule lumen

ARTERIOLAR NEPHROSLCEROSIS

Arteriolar nephrosclerosis is the manifestation of renal damage due to hypertension It is characterized by thickening and hyalinization and sometimes necrosis of the renal arterioles There is progressive loss of glomeruli and the function of the tubules to which they are attached Clinically it is characterized by proteinuria and progressive impairment of renal function which in some instances ("malignant nephrosclerosis") may be remarkably rapid Some degree of nephrosclerosis occurs in most cases of hypertension of long-standing The progressive form leading to renal insufficiency occurs in some 5 to 10% of patients with so-called "essential" hypertension if not treated with drugs that lower blood pressure

In contrast to the diffuse involvement of the kidney in *arteriolar*sclerosis (arteriolar nephrosclerosis) *arterio*sclerotic involvement of the larger vessels may lead to patchy ischemic degeneration of portions of the kidney This type of involvement is relatively benign and rarely leads to significant functional disturbance, except when, occasionally, marked stenosis of the main renal artery or its major branches produces sufficient renal ischemia to lead to hypertension

PYELONEPHRITIS

Pyelonephritis is an inflammatory disease of the kidney resulting from infection of the urinary tract with invasion of the kidney interstitium Single acute attacks may subside without significant sequelae, but repeated or long-standing infections lead to scarring and contraction of the kidney and what may be extensive loss of renal tissue The loss of renal function may be accelerated as changes due to hypertension (itself initiated by the chronic pyelonephritis) are superimposed on the results of infection and scarring Proteinuria is usually not prominent and may be absent

Disturbances of Function in Diffuse Renal Disease

PROTEINURIA (ALBUMINURIA)

Perhaps the commonest manifestation of renal disease is the presence of increased amounts of protein in the urine Normally only traces of protein are present in the urine, less than 100 mg being excreted in a 24-hr period In the presence of renal disease the amounts excreted may vary widely and reach figures of many grams Whereas protein may gain access to the urine with exudation or bleeding into any part of the urinary tract, the glomeruli are the most pertinent source of urinary protein Among the proteins of the urine, albumin predominates, but despite the fact that urinary protein is frequently referred to as albumin, the other plasma proteins are also represented However, it is true that the relative abundance of albumin is greater in the urine than the plasma

An increased concentration of protein in the glomerular filtrate, the result of an abnormal increase in the permeability of the glomerular capillaries to plasma protein, is generally accepted as the cause of proteinuria in most instances of renal disease and is one of the major features of the nephrotic syndrome, of glomerulonephritis and of arteriolar nephrosclerosis

It is worth noting that decreased reabsorption of normally filtered protein has been seriously considered as a possibly significant factor in producing proteinuria This possibility is based on the uncertainty concerning the amount of protein normally contained in the glomerular filtrate in man At a concentration of only 10 mg/100 ml, some 18 g/day might be excreted in the urine if none were

reabsorbed It may be dangerous to assume that the concentration in human glomerular filtrate is the same as that in the dog but it is worth noting that recent micropuncture studies in the dog revealed albumin concentrations in the proximal tubule averaging less than 5 mg /100 ml

On the other hand, there is good evidence not only that there is an increased filtration of protein when proteinuria is present, but that there is also an increase in the reabsorption of protein by the tubules under these conditions and that, in all probability, it is this increased protein uptake that is responsible for the histological changes in renal tubule cells in such disorders as the nephrotic syndrome

Aside from its importance as a clue to the presence of renal disease, proteinuria may, when of sufficient intensity, lead to significant depletion of the plasma proteins In the nephrotic syndrome, largely as a result of urinary losses, the plasma albumin concentration may be reduced to 20% or less of its normal value This is a significant factor in the development of nephrotic edema

The presence of proteinuria is not always a sign of significant organic disease of the kidney It may be produced in otherwise normal individuals, for example, by fever or by violent exercise Particularly likely to be misleading is the benign condition known as *orthostatic albuminuria* or *postural proteinuria* This is characterized by the occurrence, especially in young people, of proteinuria when the individual assumes the upright position The urine, after a night's rest in bed, is protein free The production of proteinuria is presumably attributable to some sort of circulatory disturbance associated with the erect posture It is important chiefly because it is frequently mistakenly interpreted as indicating the presence of latent glomerulonephritis

CYLINDRURIA

Casts are collections of protein that has precipitated in the tubules or ducts of the kidney and has been molded in the form of the structure in which it was precipitated. Cylindruria or the presence of increased number of casts in the urine is commonly associated with proteinuria of glomerular origin Aside from the diagnostic information derivable from the nature of the casts, the detection of casts is most significant as an indication that the protein in the urine has arisen from the nephrons The casts may consist of protein only (hyaline casts) or may have incorporated into them red or white blood cells, epithelial cells, etc The normal urine does not contain more than 2000 casts/24 hr

HEMATURIA

Red blood cells may be present in the urine in increased numbers (normal less than 130,000/24 hr) because of bleeding anywhere into the urinary tract Of greatest importance in the present context is that, with inflammatory disease of the glomeruli, red cells as well as protein may penetrate the glomerular capillaries Marked hematuria of renal origin is particularly characteristic of acute glomerulonephritis

IMPAIRMENT OF URINARY CONCENTRATION POLYURIA NOCTURIA

Two types of impairment of urine concentrating capacity are recognizable *1*) due to impairment of the process by which water is removed to produce urine hypertonic to plasma and *2*) due to the excretion of increased amounts of solute per functioning nephron Both types of defect contribute to the impairment of urine concentrating power observed in the presence of renal disease Specific defects in the process that renders the urine hypertonic are present in *potassium depletion* and *hypercalciuria* Since the processes responsible for rendering the urine more concentrated than plasma are located in the renal medulla, disease that disrupts the structure and/or function of this portion of the kidney, such as pyelonephritis and hydronephrosis, may be expected to have particularly marked effects on the concentrating process

On the other hand, as renal function becomes progressively impaired by any nephron-destroying disease process, impaired ability to produce concentrated urine must inevitably result Indeed, it has been shown experimentally that removal of one kidney and part of the other results in impaired ability of the remaining kidney to produce concentrated urine The situation here is quite analogous to that observed as solute excretion is markedly increased (osmotic diuresis) in the individual with normal kidneys (see previous chapter, Fig. 5 9*B*) As renal function becomes progres-

sively diminished, the rate of excretion of solute must, over the long run, remain essentially unchanged reflecting the maintenance of salt balance and the excretion of products of metabolism However, as the number of nephrons involved in urine formation becomes smaller, the amount of solute excreted by each of those remaining becomes greater in proportion Thus, the volume of isotonic fluid in each nephron reaching the site of formation of hypertonic urine is increased, and even the removal of a normal amount of water changes the concentration only to a limited extent

As the ability to form concentrated urine is impaired, the volume of urine excreted must increase The formation of increased volumes of urine is known as *polyuria*, a common manifestation of impaired renal function The production of excessive amounts of urine may cause the patient to awake at night to void, a symptom designated as *nocturia* (Of course, the ingestion of large volumes of fluid may lead to polyuria and nocturia in the absence of renal disease)

OLIGURIA ANURIA

Oliguria is the excretion of a markedly reduced volume of urine, anuria is the complete suppression of urine formation Oliguria may, at times, be physiological—that is, a normal response to dehydration and the characteristic manifestation of the formation of highly concentrated urine A particularly low urine flow is found when dehydration is accompanied by a low rate of solute excretion This combination is likely to be encountered in the individual who is excreting very little salt because of a low salt intake, because of antecedent salt loss (sweating, diarrhea, vomiting, induced diuresis) or because salt excretion is suppressed for other reasons as in the individual who is developing edema (see discussion of edema)

On the other hand, oliguria or its extreme, anuria, may be a manifestation of a severe disorder of renal function. It may occur as a complication of severe acute nephritis or as a terminal event in chronic glomerulonephritis or other chronic renal disease Most often, however, anuria or severe oliguria is a manifestation of some acute renal insult—a circumstance commonly known as *acute renal insufficiency* or *acute renal failure* Such severe acute damage may be produced by a variety of injuries and nephrotoxic substances, particularly with traumatic shock (in which circumstance it has been called the *crush syndrome*), hemorrhagic shock, incompatible blood transfusion, and mercury or carbon tetrachloride poisoning This list is far from inclusive, and acute renal insufficiency may be caused by any condition which produces prolonged hypotension and/or intense renal vasoconstriction Renal ischemia becomes a major factor in the production of acute renal insufficiency in those situations in which chemical nephrotoxins can not be implicated

The methods by which glomerular filtration is measured are not applicable in the presence of anuria, so that it is not possible to say with certainty whether, in acute renal failure, glomerular filtration ceases, or whether it continues but that complete reabsorption of the filtered fluid occurs through disrupted tubule walls Both theories have their supporters In either case, it is clear, *effective* glomerular filtration is zero

It has also been proposed that in the anuria following trauma the renal blood flow is not reduced, but actually increased, the blood being shunted away from the cortex and through the juxtamedullary and medullary regions by dilation in the latter areas This concept has been largely discredited In the presence of anuria there is a marked reduction in the renal blood flow The preferential injection of the juxtamedullary circulation which can be demonstrated under these conditions (Trueta et al , 1946) is due to the lesser involvement of this part of the renal circulation in the generally intense renal vasoconstriction

EDEMA

The accumulation of sufficient excess of extracellular fluid to produce detectable swelling of distensible tissues (edema) is a common feature of many renal diseases Since edema fluid is extracellular fluid, it has the composition of plasma except for its lower protein concentration. Like plasma, then, its major constituents are sodium chloride and water Except for localized and limited amounts of edema which can be produced by redistribution of the normal volume of extracellular fluid, the production of edema requires that an excess of both salt and water be retained within the body and, conversely, edema may be produced by any circumstance that yields a sufficient

increase of the salt and water content of the body Actually in the production of edema, it is the salt retention which is primary, water being retained to maintain the normal osmotic pressure of body fluids, and, further, in the salt, it is specifically the sodium ion which plays the predominant role, in the edematous individual, edema is increased by sodium bicarbonate (or, of course, sodium chloride) but not by ammonium or potassium chloride

In patients with renal disease, edema may be traced to three types of distrubance *1)* to *impairment of sodium excretion* due to the direct effect of disease on renal processes involved in salt excretion, *2)* to *hypoproteinemia*, and *3)* to *cardiac insufficiency*, a common complication of renal disease particularly in the presence of long-standing hypertension

1) The edema of acute glomerulonephritis is primarily of the first type, due to the effect of disease on the renal processes involved in salt excretion, salt intake exceeds the capacity of the kidney to excrete it, and edema results The predominant mechanism is probably a sudden and sharp reduction in the rate of glomerular filtration without an equivalent impairment of the capacity of the tubules to reabsorb sodium salts, a condition frequently called "glomerulotubular imbalance " Through much of the course of chronic glomerulonephritis, a new balance is struck so that, despite reduction of the rate of glomerular filtration, sodium reabsorption by the tubules is reduced to an equivalent extent, and adequate excretion of salt is maintained In the course of chronic glomerulonephritis, edema may appear as a result of any of the mechanisms noted above, because of the development of hypoproteinemia or, particularly late in the course of the disease, because glomerular filtration is reduced to such low levels that it is no longer possible to maintain more than a very low rate of salt excretion, and finally because cardiac insufficiency may be superimposed

In acute renal insufficency the capacity to excrete salt may be entirely lost Thus, any salt and fluid intake is retained in the form of edema, except to the extent to which they may be lost by extrarenal routes This situation represents the extreme of edema formation due to impairment of the renal processes for salt excretion

2) The classic example of edema resulting primarily from hypoproteinemia is that which characterizes the nephrotic syndrome In this disorder the continuous loss of large amounts of protein, particularly albumin, in the urine taxes the capacity of the body mechanisms to synthesize albumin and the plasma protein concentration falls As the plasma protein concentration is reduced, the colloid osmotic pressure is lowered, particularly since most of the colloid osmotic pressure of plasma is normally attributable to its albumin concentration The lowering of the plasma colloid osmotic pressure favors redistribution of fluid volume from plasma to interstitial space in accord with the Starling hypothesis (see sect 4, chap 12) It is important to note that although this series of events favors expansion of the interstitial fluid volume at the expense of the intravascular volume, no significant quantity of edema can be attributed to this series of events *alone*, the intravascular volume is not large enough to yield a significant volume of edema without disastrously compromising the circulation The production of hypoproteinemic edema, then, requires, as does all edema formation, retention of excess salt and water The reduced intravascular volume favored by the reduction of colloid osmotic pressure is, in all probability, the stimulus for renal salt retention mediated in part by the secretion of salt retention steroid (aldosterone) by the adrenal cortex and, probably to a major extent, by the as yet poorly understood system which controls the reabsorption of salt by the tubules A low rate of glomerular filtration may be a contributory factor

That hypoproteinemia alone is not a sufficient cause of edema in the nephrotic syndrome is indicated by the results of steroid treatment in this disorder. When fully successful, treatment with large doses of adrenocortical steroids produces disappearance of proteinuria, restoration of the plasma protein concentration, and diuresis with disappearance of the edema However, in a number of instances in which the steroid is without apparent effect on proteinuria or plasma protein concentration, withdrawal of steroid therapy is accompanied by diuresis and loss of the edema, although the plasma protein concentration remains at its previous low level Spontaneous diuresis without increase of the plasma protein concentration may also occur

3) The edema of cardiac failure, whether a complication of renal disease (hypertension) or not, depends on renal retention of excess salt

Thus, edema of any origin is a manifestation of disordered kidney function, in the case of heart failure secondary to an abnormality of cardiac function Although the sequence of events leading to edema in heart failure is disputed, the salt retention which occurs is the result of a complex of factors including increased secretion of salt-active adrenocortical steroids, some reduction of renal blood flow and glomerular filtration in many instances, and, almost certainly, the poorly understood factor or factors mentioned above The latter is manifested by the following observations The administration of salt-active steroids to normal experimental animals leads to only very little retention of sodium (much of which replaces potassium which is lost) and no edema If, however, because of borderline cardiac insufficiency or elevation of venous pressure in a major segment of the circulation, the experimental animal is prone to the development of edema, sodium will be retained and edema will form—without reduction of the rate of glomerular filtration The difference between edema formation and no edema formation in these two experimental situations is neither in glomerular filtration nor the amount of adrenocortical steroid administered, indicating involvement of the additional unidentified factors

ACIDOSIS

Inability to maintain a normal level of plasma bicarbonate (alkali reserve) in the face of an acidifying intake is a common feature of advanced renal disease of any type The most important factor responsible for the acidosis is deficient ability to form and excrete ammonia and thus maintain the plasma bicarbonate concentration in the face of the acid load provided by the acififying components of the normal diet Usually there is no defect in the ability of the kidney to render the urine acid and highly acid urines are frequently observed in even the most advanced renal failure Since there is no major change in the excretion of weak acids, the output of titratable acid is not depressed However, the reduced number of nephrons is not adequate to form and put out ammonia at a normal rate in the face of even the usual, much less any increased, demands Thus, the strong acids derived from dietary sources are excreted as the sodium or potassium, rather than the ammonium salt and the blood bicarbonate (alkali reserve) is depleted In some cases an additional factor may contribute to the acidosis in the form of a deficient ability to reabsorb bicarbonate completely when the plasma bicarbonate concentration is at normal or only slightly lowered levels

If bicarbonate is added to render the intake effectively neutral, the acidosis of renal failure may be absent In fact, if bicarbonate intake is much more than that necessary for a neutral intake, alkalosis may be relatively easily produced since the capacity to excrete bicarbonate at a rapid rate is limited as a consequence of the reduction of the rate of glomerular filtration

As opposed to the tendency to the development of acidosis, which is almost universal in advanced renal disease of any type, acidosis of renal origin may occur as the result of a more or less specific defect of renal function with renal function remaining relatively intact in other respects This relatively uncommon condition is known as *renal tubular acidosis* In contrast to the acidosis of renal insufficiency, it appears to be due specifically to an inability to render the urine normally acid and not to a deficient capacity to form ammonia In some such cases the maximum capacity to secrete hydrogen ion in exchange for sodium can be shown to be normal, since there is a normal maximum rate of bicarbonate reabsorption when bicarbonate is administered Under these conditions of bicarbonate loading, the urine is alkaline and the secretion of hydrogen ion which effects bicarbonate reabsorption occurs against only a very limited concentration gradient (see chap 1, p 5-29ff) However, as the bicarbonate concentration is reduced to normal or below, the tubules are unable to effect the complete removal of bicarbonate, this requiring that the urine be rendered highly acid Hence, bicarbonate escapes in the urine The defect appears to be in the ability to secrete hydrogen ion against any considerable concentration gradient Since the urine does not become acid, the excretion of ammonia remains low despite what is probably a normal ability to produce ammonia in the renal tubules (see p 5-31)

RENAL INSUFFICIENCY, UREMIA

Renal insufficiency is the general term for reduction of the function of the kidney It is manifested by many of those physiological disturbances already described—impaired con-

centrating capacity, polyuria, nocturia, acidosis, and a general loss of flexibility in the maintenance of normal body composition in the face of the stress of departure of the intake from ideal *Uremia* is the complex of symptoms and signs that accompany advanced renal insufficiency It is characterized by anemia, nausea, vomiting, weakness, twitching, frequently convulsions, and, terminally, stupor and coma Although it derives its name from the elevated blood urea concentration, which is a universal finding in the presence of the syndrome, the urea retention is not primarily responsible for the symptoms

That urea itself is not responsible for the abnormalities associated with uremia is shown by several observations Urea may be administered to normal individuals without producing significant disturbances In the uremic individual, the blood urea can be considerably reduced by restriction of the protein intake without otherwise affecting the course of the disorder The improvement which can be produced in the uremic patient by dialysis with the "artificial kidney" is not lost when urea is included in the dialysis bath so that the blood urea is not removed by the dialysis

Disturbances of acid-base balance and fluid and electrolyte metabolism are also commonly present in uremia, and may, at times, contribute to or be responsible for certain of the symptoms Nevertheless, these electrolyte imbalances can not be considered to be the major contributor to the picture of uremia, and the factor or factors primarily responsible for the symptoms and signs remains unidentified It is generally believed that noxious products of metabolism may, in the presence of depressed renal function, accumulate in the body fluids till toxic concentrations are reached These substances, if they exist, have never been adequately identified

INTRINSIC DEFECTS OF RENAL TUBULAR FUNCTION

As opposed to the disturbances of renal function attributable to diffuse inflammatory or degenerative kidney disease, there are a number of abnormalities due to deficiency of certain transport mechanisms or groups of transport mechanisms Many of these defects are familial or congenital though some may be acquired They are for the most part uncommon, but warrant brief mention because of their unusual physiological interest

One such defect, *renal tubular acidosis*, has been mentioned above and the impairment of sodium-hydrogen ion exchange described The disorder is, at least in some instances, familial It is frequently associated with nephrocalcinosis and pyelonephritis, but these are quite probably secondary rather than the cause of the renal abnormality

Nephrogenic diabetes insipidus is a rare congenital familial disorder most often encountered in males and characterized by the continuous obligatory excretion of dilute urine, and consequently intense polyuria and secondarily polydipsia The abnormality is apparently due to inability of the distal portions of the renal tubules to respond to pituitary antidiuretic hormone by an increase in the permeability to water Thus, the urine, diluted by salt removal in the distal portions of the nephron, fails to give up the excess of water in the tubules, and the water is lost in the urine True or *pituitary diabetes insipidus* is characterized by the same symptoms and signs but is due to a lack of antidiuretic hormone secretion The tubules are normally responsive to antidiuretic hormone in this disorder, which is much more frequent than the nephrogenic type Pituitary diabetes insipidus is usually an acquired abnormality secondary to injury to the pituitary and/or hypothalamus

Renal glycosuria is characterized by the excretion of glucose in the urine when the concentration in the plasma is normal or low The abnormality is a defect in the mechanism by which glucose is reabsorbed in the proximal tubule, other functions of the proximal tubule may be normal The specific nature of the derangement is not known There is some depression of the maximum capacity to reabsorb glucose (Tm_G), but there is also considerable excretion of glucose when the concentration is below that required to saturate the reabsorptive mechanism The possible interpretations of this "splay in the titration curve" are discussed on page 5-12 of the preceding chapter

Renal Hypophosphatemia (or Vitamin D resistant rickets) is a familial disorder characterized by excretion of phosphate in the urine when the plasma concentration is low It is attributable to a defect of the phosphate reabsorptive mechanism analogous to that for glucose in renal glycosuria

Amino-aciduria. Defects of amino acid reabsorption occur as specific defects of certain transport mechanisms and as part of more generalized disorders of renal tubular function An example of the specific type is that known as *essential cystinuria* in which the abnormality is limited to diminished reabsorption of cystine, lysine, arginine and ornithine The chief hazard derives from the formation of stones from the poorly soluble cystine Abnormalities involving certain other specific amino acids or groups of amino acids are also known In addition

defects of amino acid reabsorption may occur in several other disorders including, among others, Wilson's disease (fundamentally a disturbance of copper metabolism), heavy metal poisoning, and the Fanconi syndrome in which amino-aciduria is associated with defective reabsorption of glucose and phosphate and sometimes with renal tubular acidosis

Disturbances of Renal Function of Extrarenal Origin

The function of the kidney is dependent, for the maintenance of glomerular filtration, on the adequacy of the circulation The effect of shock, leading, as it may, to sustained damage to the kidney, has already been discussed In addition several of the renal transport systems for electrolytes and water are regulated by the activity of endocrine organs The function of these systems is therefore subject to disturbance through abnormal activity of the endocrine organs in the form of either deficient or excessive secretion of hormones acting on the kidney Two such abnormalities have already been described in relation to intrinsic renal disorders which they resemble or to which they are related

1) *Pituitary diabetes insipidus* in which, because of deficiency of antidiuretic hormone secretion, renal water reabsorption is diminished and a dilute urine is continuously formed The converse of this disorder may occur and a concentrated urine be formed when a reduced osmotic pressure of body fluids would ordinarily lead to the formation of dilute urine, most often this is due to the induction of antidiuretic hormone secretion by nonosmotic stimuli, occasionally to ectopic secretion of vasopressin by tumors

2) The excessive reabsorption of salt that characterizes *edema formation* and the contribution of hypersecretion of aldosterone have been discussed in relation to edema Excessive secretion of aldosterone leads to accumulation of edema only in the presence of factors predisposing to edema formation In the absence of such factors, the increased sodium reabsorption is limited, and the major manifestation of hypersecretion of aldosterone is *potassium loss* resulting from stimulation of secretion of potassium ion in exchange for reabsorbed sodium If potassium loss due to intrinsic renal disease with defective potassium reabsorption exists, it has not been demonstrated, and excessive renal potassium excretion appears to be due to excessive secretion of potassium by the renal tubules, generally stimulated by excess adrenocortical hormone

In the absence of adrenocortical hormones (*adrenal insufficiency*, Addison's disease) reabsorption of sodium and secretion of potassium by the renal tubules are defective Sodium loss in the urine proceeds despite contraction of the extracellular fluid volume, which may become markedly reduced The potassium concentration of body fluids becomes elevated Abnormal loss of sodium in the urine despite sodium depletion may occasionally result from renal disease, and may be produced by inhibitors of sodium reabsorption (diuretics)

The reabsorption of phosphate by the renal tubules is regulated by parathyroid hormone Excessive secretion of parathyroid hormone (*hyperparathyroidism*) is associated with loss of phosphate in the urine, loss of parathyroid function (*hypoparathyroidism*) leads to elevation of the plasma phosphate concentration without an equivalent increase in urinary excretion

THE URINARY BLADDER

ANATOMIC CONSIDERATIONS

Urine leaving the tips of the renal papillae passes through three functionally distinct divisions of the urinary tract in its course to the exterior The first of these, termed the upper urinary tract, is made up of the minor and major calyces of the two kidneys, the renal pelves, and the ureters It functions as a low-pressure, actively propulsive conduit with a small, fixed capacity, transmitting urine rapidly to the urinary bladder by peristaltic wave motion The second division, the urinary bladder, acts as a reservoir, storing urine at low pressures and periodically discharging it at much higher pressures The third division, comprising the male urethra below the urogenital diaphragm, serves as a passive conduit for the urinary stream and, except for the intrinsic resistance offered to the flow of fluid, plays no part in the mechanics of voiding This structure has no homologue in the female, a fact which underscores its role as an essential male genital organ, associated anatomically rather than functionally with the urinary tract

The entire urinary tract is lined by transitional epithelium as far distally as the membranous urethra This epithelium, 5 to 7 cells thick when the urinary organs are empty, rests on a fibroelastic lamina propria which is com-

monly but incorrectly designated the submucosa (A muscularis mucosae is not present in the urinary system, and, therefore, a true submucous coat which by definition lies deep to this muscular layer cannot be present The term "submucosa" will nevertheless be retained, since it is honored by long and extensive usage) Transitional cells have the ability to slide over one another, thereby increasing their surface area at the expense of the thickness of the cell layer without breaking the continuity of the epithelium Deep mucosal folds or rugae are present in the collapsed ureter and bladder, which, together with the unique properties of the epithelial coat, permit the considerable dilatation which these structures undergo in the course of urine transportation and storage The submucosa rests on a smooth muscle coat, the muscularis, and is in turn surrounded by a loose areolar adventitia which in no way limits distensibility

Anatomic adaptation to function resides chiefly in the muscularis of the urinary tract It was formerly believed that the smooth muscle of the upper tract and bladder was disposed in three layers of unvarying thickness, an inner longitudinal, a middle circular, and an outer longitudinal This concept is not supported by anatomic microdissections, and fails to explain the pattern of motility of these organs The muscularis of the upper urinary tract is now believed to be arranged as interlacing bands of smooth muscle fibres, spiralling in both a clockwise and counterclockwise direction from the necks of the minor calyces to the trigone of the bladder In the calyces and renal pelvis, the majority of spirals have a shallow angle of ascent, and the majority of the muscle in this area appears circular Near the ureteropelvic junction and in the upper few centimeters of the ureter, the spirals wind much less sharply and the muscle bundles appear predominantly longitudinal The major length of the ureter is again supported mainly by tightly wound spirals with only a few longitudinally arranged bundles, but, in the lowermost portion of the ureter, where it passes through the bladder wall, the muscular spirals disappear and the intramural ureter is surrounded only by a thin coat of entirely longitudinal muscle These fibers are continued beyond the ureteral orifices, where they fan out and mingle with their opposite numbers to form the bulk of the trigonal muscle

The individual smooth muscle cells of the upper urinary tract are linked by intercellular bridges, forming a syncytium capable of conducting an action potential without the necessity of extrinsic nervous mechanisms The force and direction of the resulting muscular contraction depends on the predominant directional arrangement of the muscle bundles at any given level

The lower ends of the ureters enter the posterolateral angles of the bladder wall and, pursuing an oblique course, first through an hiatus in the vesical muscle, and then within the vesical submucosa, terminate in orifices at the lateral angles of the trigone The total length of ureter enclosed by the bladder in the adult is approximately 2 cm , with an average diameter of 3 5 mm Terminally, the ureter is anchored firmly to the trigone by continuation of its longitudinal muscular coat, and by fusion of its roof with the mucous membrane of the bladder Where is passes through the ureteral hiatus, however, there is only the loosest of areolar connections between vesical and ureteral muscle (Waldeyer's space) in order that the bladder may "slide up" around the ureter as it fills with urine Externally the ureter is invested by the reflection of fascicles of the detrusor musculature onto the lower several centimeters (Fig 5 14) There is thus no anatomical basis of a circular muscular sphincter in this area, known as the ureterovesical junction, which might prevent reflux of urine from the bladder to the upper urinary tract The ureterovesical junction nevertheless remains closed except when urine is flowing into the bladder Reflux is prevented by the oblique course of the ureter through the bladder wall, which is in turn maintained by the firm terminal anchoring of the ureter to the trigone of the

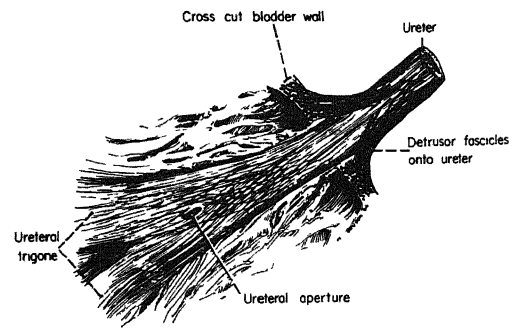


Fig 5 14 Reflectance of fascicles of muscle from bladder onto lower portion of ureter Source Woodburne, 1964

bladder and the firm support of its posterior wall by the detrusor muscle of the bladder When intravesical pressure exceeds ureteral pressure, as in voiding, the flexible anterior wall of the submucosal ureter is collapsed against the firm posterior muscular supports while fixation to the trigone maintains its length The resistance to flow through a tube of zero diameter is infinitely high, and reflux cannot occur

The bladder, when distended, consists of a spherical portion called the fundus and a cylindrical portion called the posterior urethra (Fig 5 15) The male posterior urethra includes the prostatic and membranous urethra, and in the female it includes the entire proximal three-fourths of the urethra The fundus and posterior urethra should be considered as one unit, both anatomically and functionally, as both are derived embryologically from the urogenital portion of the cloaca

The muscularis body of the urinary bladder, known as the detrusor muscle, is composed of an interwoven network of decussating smooth muscle bundles, coursing in random fashion in all directions, and at all depths in the bladder wall There is no evidence for the three concentric layers of fibers formerly believed to be present At the vesical neck and in the posterior urethra the orientation of these muscle bundles is primarily longitudinal To stress the longitudinal orientation of the smooth muscle in the posterior urethra is not to deny that other components of the bladder wall continue into the urethra in spirally and circular directions external to the inner longitudinal muscle A true annular sphincter has never been demonstrated at the vesical neck, although there is

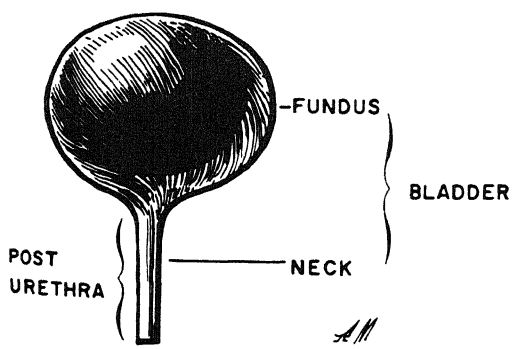


Fig 5 15 The bladder when distended consists of a spherical portion called the fundus and a cylindrical portion called the posterior urethra Source Lapides 1970

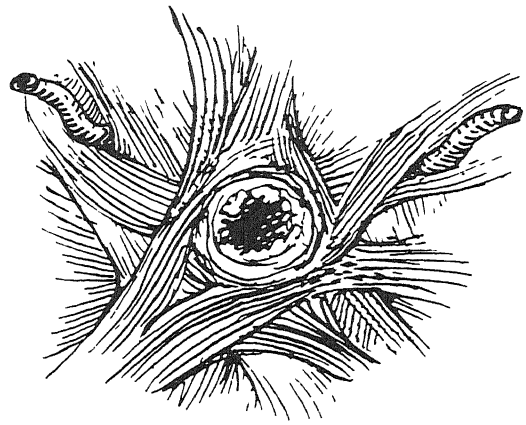


Fig 5 16 Dissection of musculature at vesical neck of bladder Source Woodburne 1960

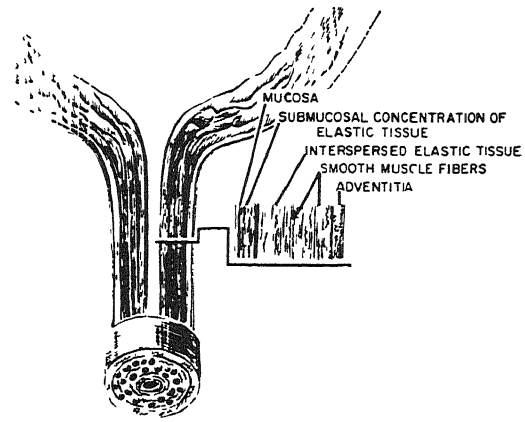


Fig 5 17 Disposition of elastic tissue in posterior urethra Source Lapides 1970

a minor accumulation of muscle fibers in this area In dissection of the vesical neck it has been noted that fascicles of smooth muscle arch toward and then away from the vesical neck and their pull would be in a diverging rather than in a constricting motion, hence, constituting an opening mechanism (Fig 5 16)

In addition there is a considerable amount of elastic tissue found in the bladder and in the posterior urethra The amount of elastic tissue is most prominent at the vesical neck and along the posterior urethra This elastic tissue is in the submucosa and intermingled with the muscle bundles (Fig 5 17)

The structure of the external sphincter and its role in voluntary compression of the urethra has been briefly considered Urethral compression is also exerted to a minor extent by the levator ani muscles, and the small muscles of

the perineum, notably the bulbocavernosus, which is also capable of stripping the male urethra following micturition

Neuroanatomic Considerations

THE UPPER URINARY TRACT

Nerve fibers and ganglion cells are present in the adventitia, muscular and submucosal coats of the ureter, originating from the sympathetic and parasympathetic divisions of the autonomic nervous system The presence of this innervation is unnecessary for the maintenance of muscular tone, or for the initiation, maintenance and variation of pelviureteric contractions. Drugs affecting the autonomic nervous system at all synaptic levels have no effect on ureteral motility in pharmacologic doses, whether administered systemically or applied directly to the ureteral wall Ureteral denervation, likewise, has no effect The best available evidence indicates that ureteral motility is an autonomous function, dependent on normal ureteral metabolism, and coordinated via the intercellular smooth muscle bridges of the muscularis The stimulus to contraction is stretching of the muscle fibres The ureter is richly supplied with visceral afferent fibres terminating in the L1 and T12 segments of the spinal cord Acute ureteral distension is the adequate stimulus, giving rise to ureteral pain, referred along the corresponding cutaneous dermatomes This pain may be abolished by section of the ureteral nerves without interfering in any way with peristalsis or tone

BLADDER BLADDER NECK AND EXTERNAL SPHINCTER

These structures function as a unit in the maintainance of continence and the act of micturition Their nerve supply is, therefore, functionally integrated, and can be logically considered under a single heading Innervation is both motor and sensory, derived from the somatic and autonomic nervous systems, with integration on both segmental reflex and suprasegmental levels

Motor Innervation Sympathetic white rami arising from the lowest thoracic and upper two lumbar segments of the intermediolateral column of grey matter of the spinal cord pass through the sympathetic chain to synapse in the celiac and mesenteric ganglia, from which grey rami flow into the pelvis as part of the superior hypogastric plexus (presacral nerve) At the promontory of the sacrum the plexus bifurcates, to form the right and left inferior hypogastric plexi (hypogastric nerves) which run deep into the pelvis to form the pelvic plexus From here, fibers pass to smaller plexi surrounding the bladder, prostate and posterior urethra These efferents, contrary to previous teaching, have no known effect on vesical and bladder neck function, through their action on the trigone and urethral crest, however, they may close the bladder neck during ejaculation

Parasympathetic motor fibers to bladder and bladder neck arise in the intermediolateral column in the second, third and fourth segments of the sacral cord, and travel as medullated fibers in the pelvic nerves (*N erigentes*) to the pelvic plexus Synapses occur with postganglionic neurons either here, or with ganglion cells within the bladder wall, with grey rami taking a short, direct course to the vesical muscle The parasympathetic division is the sole motor supply of the detrusor, both fundus and posterior urethra The time-honored theory of dual antagonistic innervation of the body and neck of the bladder, in which sympathetic impulses relaxed the detrusor and constricted the internal sphincter (syn = bladder neck), and parasympathetic impulses produced the reverse, is not supported by experimental evidence Presacral nerve section in the human has no effect on micturition, nor do sympathomimetic drugs Bilateral pelvic nerve section causes complete detrusor paralysis, with urinary retention and overflow incontinence, while their stimulation causes immediate detrusor contraction

The somatic motor supply of the striated muscles of the perineum, including the external sphincter, is carried via the pudendal nerves from the anterior horn cells of sacral segments 2, 3, and 4 These fibers maintain tonic contraction of the external sphincter except when reflexly inhibited during micturition Their bilateral section causes total relaxation and degeneration of the external sphincter muscle Continence is not affected, but rather, the ability is lost rapidly and voluntarily to arrest the urinary stream, once initiated.

Sensory Innervation. Afferent fibers, both visceral and somatic, follow the three routes taken by the efferent fibers of bladder, urethra, and perineum

Some sensory fibers arising from bladder and posterior urethra follow sympathetic efferent pathways to cell stations in the dorsal root ganglia The lower limit of their central connections is at L2, while their upper limit has been variously placed from T11 to as high as T6, following clinical observations on patients with transections of the spinal cord The exact role of these fibers is obscure They probably carry pain and some proprioceptive and temperature sensation, but their section produces no significant deficit They are said to account for the vague, often painful sensations due to vesical overdistension and spasm experienced by paraplegics with spinal cord transections below T12

All the important afferent fibers of the bladder, and posterior urethra, especially those associated with sensation of fullness and desire to void but which are also useful thresholds of pain and temperature sensation, follow the pelvic nerves to the dorsal root ganglia of S2–4 Section of these nerves results in the loss of all useful vesical sensation

The pudendal nerves carry superficial and deep somatic sensation from the distal bladder neck, and adjacent perineal structures

Reflex Centers and Suprasegmental-Pathways. Centers for the control and integration of micturition are situated in sacral spinal cord, brainstem, and cerebral cortex

A spinal center for the reflex control of micturition exists in the second, third, and fourth segments of the spinal cord (Fig 5 18) Two

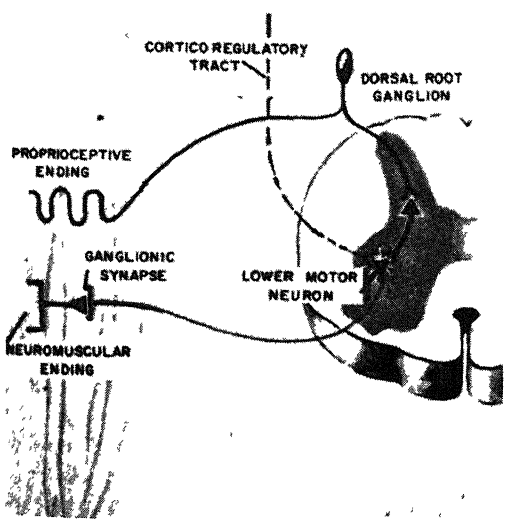


Fig 5 18 Spinal reflex arc in sacral cord concerned with micturition Source Lapides 1970

important reflexes are served Visceral afferent impulses from stretch receptors in the bladder wall travel via the pelvic nerves to the micturition center, which, unless inhibited by higher centers, sends efferent impulses via the pelvic nerves resulting in detrusor contraction This has the effect of forcing urine into the posterior urethra, initiating a second reflex Visceral afferent endings in the posterior urethra, stimulated by the presence of urine, send impulses via the pelvic nerves to the spinal micturition center, which transmits inhibitory impulses to the anterior horn cells of S2–4, decreasing the frequency of impulses in the pudendal nerve, and thereby causing relaxation of the external sphincter

The activity of the spinal micturition center is modified by a higher reflex center in the brainstem, and by cortical activity under the control of the will Pain and temperature afferents reach these centers via the crossed spinothalamic tracts after synapse in the dorsal horn cells at the appropriate segmental level Proprioceptive sensations, especially the desire to void and the feeling of fullness, turn into the ipsilateral funiculus gracilis without synapse, reach the medulla where they effect synapse in the nucleus gracilis, and then cross the midline to join the spinothalamic tracts and follow the great sensory pathways to the brainstem and cortical centers for micturition Regulatory pathways descend from these centers, via the crossed and uncrossed pyramidal and extrapyramidal tracts, to modify the activity of the spinal micturition center

The location of the brainstem center for micturition is uncertain in man, but probably resides in the pons It exerts both facilitory and inhibitory effects of the spinal center It is responsible for sustaining detrusor contraction and external sphincter relaxation until the bladder is completely emptied, this is necessary because, after partial emptying of the bladder, proprioceptive afferent impulses to the spinal center fall below the threshold necessary to maintain the micturition reflex The brainstem probably also inhibits the spinal micturition reflex at an unconscious level, even when the bladder is adequately filled

The cerebral cortex receives vesical representation at the extreme medial aspect of the sensory-motor strip (paracental lobule) Here the degree of fullness of the bladder, the slight burning felt on initiating voiding, and the pain of vesical spasm are consciously appreciated

Here too, voluntary control is exerted over the lower micturition centers Thus, although micturition can proceed normally in the absence of cortical control, the cortex inhibits lower centers until a suitable time and place for voiding can be found Voiding is then allowed to proceed automatically by removing cortical inhibition Micturition can also be voluntarily initiated by the cerebral cortex in the absence of a desire to void or, once begun can be interrupted at any time

Mechanical Considerations

The urinary system is a series of pipes and reservoirs transmitting and storing fluid, and as such is subject to the laws of fluid mechanics Most important of these is LaPlace's law, which states

$$P = 2T/R \text{ in a sphere,}$$

and

$$P = T/R \text{ in a tube,}$$

where

P = intraluminal pressure (sphere), resistance to flow (tube), (provided external pressure is constant)

T = tension on the wall of the sphere or tube (i e , the sum of the force of active tension developed by muscular contraction plus the passive tension exerted by the elastic properties of the wall)

R = radius of the tube or sphere

These relationships hold good, regardless of the characteristics of the wall of the viscus, provided the system is at equilibrium

Certain corollaries of this law provide a mechanical explanation for the behavior of the upper urinary tract and bladder, and will be emphasized in the appropriate sections below One example will be recalled, namely, the absence of reflux when the intravesical ureter is collapsed Under these conditions, the radius (R) is very small, so that for any given tension on the ureteral wall, the resistance to flow (P) in the ureter is extremely high, certainly higher than the intravesical pressure

Urinary Transport in the Upper Urinary Tract

Distension is the adequate stimulus to contraction throughout the entire upper urinary tract. Urine flowing into the minor calyces creates distension, triggering a calyceal contraction which, because of the predominantly circular muscle in the area, causes narrowing of the calyx and forces urine into the renal pelvis Calyceal contractions occur independently of one another and of renal pelvic contractions Calyceal pressures follow renal pelvic pressures, because the two are in free communication at all times

The renal pelvis is relatively indistensible beyond its fairly low capacity of 3 to 5 ml Like the calyces it undergoes rhythmic dilatation and contraction, referred to as diastole and systole, respectively It fills passively from the calyces under very low pressures during diastole, and then contracts, expressing only a small percentage of its volume into the upper end of the ureter The normal intrapelvic pressure seldom exceeds more than a few centimeters of water, even in systole, but tends to be higher in the narrow, ureter-like pelvis than in the large, ampullary-type pelvis—a direct result of LaPlace's law

Efforts to locate a pelviureteric pacemaker have failed, but the contraction wave begins somewhere in the region of the pelvicalyceal junction and sweeps distally at a relatively constant rate of about 3 cm /sec The pelvis is comprised of both circular and longitudinal muscle, and in systole, therefore, it decreases in volume and approximates itself toward the kidney

The upper end of the ureter is in communication with, and fills with, the renal pelvis during diastole, the ureteropelvic junction remaining open during this phase, and the pelvis and upper together forming the functional renal pelvis Contraction of the renal pelvis and closure of the ureteropelvic junction occur almost simultaneously, trapping urine within the renal pelvis, above, and within the uppermost ureter below the ureteropelvic junction The ureteropelvic junction may be defined, physiologically, as the most proximal point of closure of the ureter during systole of the renal pelvis, and its localization varies somewhat with the volume of urine flow

The ureteral cone is the segment of ureter in open communication with the renal pelvis during diastole Its proximal limit is the pelviureteric junction, but its size, shape, and distal limit depends on the rate of urine flow and the body position, filling being reduced in the supine position

The frequency of pelviureteric contractions ranges from 1-5/min at normal rates of flow, tending to occur at regular intervals, although at different rates on the two sides Sudden reduction of urine formation may or may not result in a transient decrease in the frequency of contractions, following which the rate returns to its normal frequency, even in the presence of complete anuria Higher than normal rates of urine flow increase the amplitude of contractions, provided the ureteral cone is capable of coping with the increased volumes If, however, renal pelvic distension occurs, the frequency of contractions increases At the highest flow rates, the entire pelvis and ureter remain open, urine transport being a function of the pelvivesical pressure gradient

The closure of the pelviureteric junction marks the origin of a peristaltic wave which sweeps down the ureter, transporting the urine trapped in the ureteral cone ahead of it A single peristaltic wave rarely strips the ureter of urine completely, but leaves behind from one to four tapered columns of urine of varying length, called urine spindles, separated by lengths of contracted, empty ureter Spindles tend to be formed under conditions causing resistance to urine flow, such as recumbency, high urine flow rates, and high intravesical pressure, which lead to retention of urine in the upper tract Small spindles fill from above and empty from below as the peristaltic wave sweeps over them, but large spindles can contract as a unit, rapidly expelling urine toward the bladder Spindle formation is a more efficient method of emptying the ureter than peristalsis, the transition from the latter to the former method of urine propulsion occurs smoothly, and is a function of the volume retained in the upper tract

Circular muscle is absent in the intravesical ureter, which, therefore, plays a passive role in urine transport Urine is forced in intermittent jets through this portion of the upper tract, ahead of the contraction waves descending from above Just as the spurt of urine ceases, the longitudinal muscle of the submucosal ureter contracts, drawing the lateral angle of the trigone slightly upwards and outwards

Pressures in the ureter rise progressively as the bladder is approached being normally about 15 cm of water in the upper ureter, and 30 cm in the pelvic ureter during systole These pressures are not transmitted to the kidney as long as the ureter distal to the ureteral cone remains closed, and are quite sufficient to maintain a ureterovesical pressure gradient under normal conditions

Vesical Physiology

FILLING OF THE BLADDER

The normal bladder is completely empty at the end of micturition, the intravesical pressure is equal to the intra-abdominal pressure As the bladder fills, it adjusts its tone to its capacity, so that minimal alterations of intravesical pressure occur over a wide range of intravesical volumes This phenomenon, sometimes called adaptation, is an inherent property of the smooth muscle of the detrusor, and is not dependent on nervous mechanisms It is, therefore, present in denervated bladders, but not after death The adaptation is not to a constant pressure, but always to a pressure slightly higher than existed before the addition of the fluid increment

In the adult, direct measurements show a gradual smooth increase in intravesical pressure to about 10 cm of water at a capacity of approximately 400 ml (Fig 5 19) The first sensation of bladder filling generally occurs at a capacity of 100 to 150 ml , and the first desire to void at between 150 and 250 ml The physiological capacity of the bladder, that is, the maximum intravesical volume tolerated without undue discomfort and at which micturition is normally effected, ranges between 250 and 450 ml

Filling of the bladder beyond its physiological capacity is associated with progressive

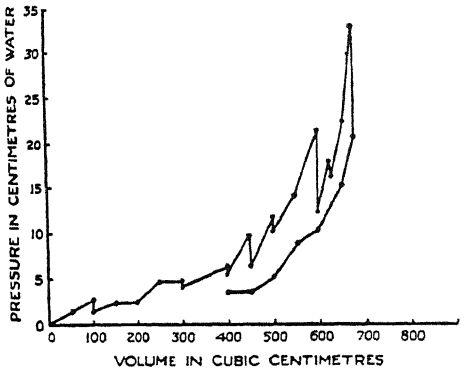


Fig 5 19 Curve of pressure changes in human bladder during filling and emptying Involuntary contractions and sharply rising pressures above 450 to 500 ml indicate filling beyond physiologic capacity

failure of adaptation, and intravesical pressure rises much more sharply relative to the increase in intravesical volume Despite strong cortical inhibition, efferent impulses reach the detrusor, causing involuntary contractions At first these are of low pressure and not associated with conscious sensation, but, as filling continues, they are associated with sharp rises in pressure, and felt consciously as an acute urge to void

The physiological bladder capacity varies greatly with age and psychic factors At birth it ranges from 20 to 50 ml, increasing four times during the 1st year of life In the adult it can be as high as 600 ml, especially in persons whose occupations make access to lavatory facilities difficult In all cases, the capacity is about twice that at which the first desire to void is felt The anatomic capacity of the bladder—the capacity just before rupture occurs—is well above 1 liter, and is never approached under physiological conditions

CONTINENCE

In addition to accepting urine from the ureters, and periodically discharging urine under voluntary control, the bladder must also retain urine As the bladder fills with urine, continence is maintained by the proximal 3 cm of the urethra This is accomplished by the continuous tone of the smooth muscle augmented by the rich elastic tissue in this area, both mechanisms tending to compress the lumen of the urethra The urinary sphincter is thus a tube of smooth muscle and elastic tissue with its greatest resistance in the mid urethra and least resistance at the vesical neck (Fig 5 20)

The external sphincter is not normally necessary to continence, and its paralysis does not cause leakage of urine, provided the posterior urethra is intact Under conditions of sudden increase of intra-abdominal pressure, such as occur with coughing, intravesical pressure may reach 50 to 80 cm of water, forcing urine into the bladder neck Leakage is prevented under these circumstances by reflex contraction of the external urethral sphincter and levator ani muscles

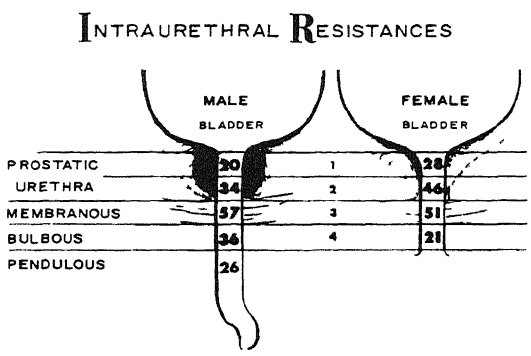


Fig 5 20 Intraurethral pressures in the male and female posterior urethra Source Lapides 1970

MICTURITION

Micturition may be defined as the periodic complete emptying of the urinary bladder, under voluntary control except in infancy The nervous, muscular and mechanical factors essential to the act have all been discussed, and must now be integrated

At the end of voiding, the bladder is empty, the bladder neck is closed, and the intravesical pressure is close to zero The bladder slowly fills as ureteral peristalsis forces jets of urine through the aperistaltic intramural ureter Despite a rising intravesical volume, the intravesical pressure remains low, due to the phenomenon of accommodation inherent in the smooth muscle of the vesical wall, and at no time does it exceed either the intraureteral pressure at the ureterovesical junction, or the urethral resistance provided by the closed posterior urethra

As the bladder radius increases, the tension on the vesical wall increases, even though the intravesical pressure remains almost constant (Laplace's law) Tension is the adequate stimulus for the proprioceptive end organs in the vesical wall which send visceral afferent impulses via the pelvic nerves to terminate in the spinal reflex center for micturition (S2-4), and also in the fasciculus gracilis to the reflex center for micturition in the brain step (pons?) and in the voluntary center for micturition in the paracentral lobule of cerebral cortex The first feeling of filling of the bladder is generally appreciated at a capacity of 100 to 150 ml, and the first desire to void, accompanied by a vague feeling in the penis or perineum, at 150 to 250 ml At this stage, the spinal reflex center for micturition is inhibited by impulses from the brainstem center via the corticoregulatory tracts, preventing efferent impulses from leaving the spinal center If micturition is unduly postponed, a feeling of fullness, then discomfort, and finally pain results The latter

sensation is transmitted to the cortex via the lateral spinothalamic tracts Under these conditions, voiding is inhibited by cortical impulses blocking the discharge of the spinal center

When an opportunity is found to empty the bladder, the voluntary restraint placed upon the spinal micturition center is lifted, and micturition proceeds automatically The cessation of inhibitory impulses from the cerebral cortex (in the case of a full bladder), or the facilatory impulses from the cortex (in the case of low intravesical volume), permit the discharge of parasympathetic efferent impulses along the pelvic nerves to synapses in the pelvic plexi and bladder wall, and thence via postganglionic rami to the body of the detrusor muscle and the posterior urethra A voiding contraction of the detrusor results, generating intravesical pressures of 50 to 150 cm of water At the same time, the detrusor contraction shortens and widens the posterior urethra (Fig 5 21) The effect is to increase intravesical pressure and decrease urethral resistance, and a bolus of urine is forced into the posterior urethra Urine in the posterior urethra stimulates parasympathetic visceral afferents which send impulses via the pelvic nerves to the spinal center This reflexly relaxes the external sphincter by decreasing the frequency of efferent impulses which normally travel from the anterior horn cells of S2–4 via the pudendal nerves to keep the external sphincter in a state of tonic contraction Urine flows through the urethra and, in the male, through a passive conduit, the penile urethra, to the exterior

Certain muscular movements not essential to micturition commonly accompany the act At the onset of micturition the levatores ani and perineal muscles are relaxed, thereby shortening the posterior urethra and decreasing urethral resistance At the same time the glottis closes, the diaphragm descends, and the abdominal wall muscles contract, accelerating the flow of urine by raising intra-abdominal pressure which in turn secondarily increases intravesical pressure

A voiding contraction, once initiated, is normally maintained until all the urine has been discharged from the urinary bladder This is a function of facilitory impulses from the brainstem The bladder contracts in all directions like a toy balloon deflating from its neck When the bladder and posterior urethra have emptied, the external sphincter closes, the detrusor relaxes, and finally the posterior urethra closes gradually

Arrest of the voiding stream, once initiated, is accomplished by a powerful voluntary contraction of the external sphincter and perineal muscles, and is associated with a sense of burning and urgency due to stimulation of somatic afferents in the region of the membranous urethra If the contraction is sustained, urine is forced backwards into the bladder by collapse of the posterior urethra, associated with gradual relaxation of the detrusor muscle

Micturition begins in the 5th intrauterine month with the onset of urinary secretion It remains a reflex act until approximately 2½ years of age, at which time it begins to come under cortical control Complete continence is usually achieved by the age of three

Uninhibited contractions of the bladder occur in young children up to a variable age, and may be attributed to an underdeveloped control of the spinal reflex arc by the upper brain center that control micturition in the adult

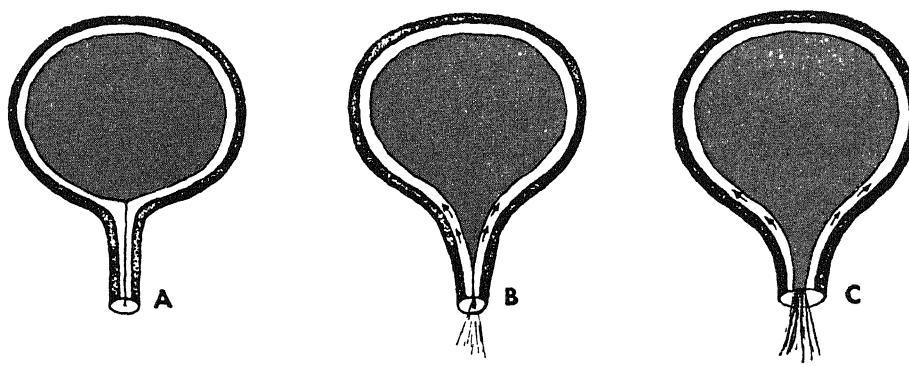


Fig 5 21 The bladder neck in continence and micturition Detrusor contraction decreases urethral resistance by shortening and widening the bladder neck Source Urology ed M F Campbell 1963 (See Lapides 1970)

BIBLIOGRAPHY

ANDERSEN B AND H H USSING Solvent drag on nonelectrolytes during osmotic flow through isolated toad skin and its response to antidiuretic hormone *Acta Physiol Scand* 39 228-239 1957

BENNETT C M B M BRENNER AND R W BERLINER Micro puncture study of nephron function in the Rhesus monkey *J Clin Invest* 47 203-216 1968

BERGER E Y S J FARBER AND D P EARLE JR Renal excretion of mannitol *Proc Soc Exp Biol Med* 66 62-66 1947

BERLINER R W Renal mechanisms for potassium excretion *Harvey Lect* 55 141-171 1961

BERLINER R W AND C M BENNETT Concentration of Urine in the mammalian kidney *Amer J Med* 42 777-789 1967

BEYER K H, L D WRIGHT H R SKEGGS H F RUSSO AND G A SHANER Renal clearance of essential amino acids their competition for reabsorption by the renal tubules *Amer J Physiol* 151 202-210 1947

BOWMAN W On the structure and use of the Malpighian bodies of the kidney with observations on the circulation through that gland *Phil Trans Roy Soc* 97-80 1842

BRAZEAU P AND A GILMAN Effect of plasma CO tension on renal tubular reabsorption of bicarbonate *Amer J Physiol* 175 33-38 1953

CHASIS H, H A RANGES, W GOLDRING, AND H W SMITH The control of renal blood flow and glomerular filtration in normal man *J Clin Invest* 17 683-697 1938

CLAPP J R AND R R ROBINSON Osmolality of distal tubular fluid in the dog *J Clin Invest* 45 1847-1853 1966

CUSHNY A R The secretion of urine Ed London Longmans, Green 1926

DAVIDSON D G N G LEVINSKY AND R W BERLINER Maintenance of potassium excretion despite reduction of glomerular filtration during sodium diuresis *J Clin Invest* 37 548-555, 1958

DAVIES B M A AND J YUDKIN Studies in biochemical adaptation The origin of urinary ammonia as indicated by the effect of chronic acidosis and alkalosis on some renal enzymes in the rat *Biochem J* 52 407-412, 1952

DE WARDENER, H E, I H MILLS W F CLAPHAM, AND C J HAYTER Studies on the efferent mechanism of the sodium diuresis which follows the administration of intravenous saline in the dog *Clin Sci* 21 249-258 1961

FINDLEY, T, JR AND H L WHITE The response of normal individuals and patients with diabetes insipidus to the ingestion of water *J Clin Invest* 16 197-202 1937

GAMBLE J L C F McKHANN A M BUTLER AND E TUTHILL An economy of water in renal function referable to urea *Amer J Physiol* 109 139-154, 1934

GOTTSCHALK, C W Osmotic concentration and dilution of urine *Amer J Med* 36 670-685 1964

GOTTSCHALK C W AND M MYLLE Micropuncture study of the mammalian urinary concentrating mechanism evidence for the countercurrent hypothesis *Amer J Physiol* 196 927-936 1959

GRANTHAM J J, AND M B BURG Effect of vasopressin and cyclic AMP on permeability of isolated collecting tubules *Amer J Physiol* 211 255-259, 1966

GUTMAN, A B T F YÜ, AND L BERGER Tubular secretion of urate in man *J Clin Invest* 38 1778-1781, 1959

HEIDENHAIN R Versuche uber den Vorgang der Harnabsonderung *Arch Ges Physiol* 9 1-27, 1874

HÖBER, R Effect of some sulfanilamides on renal secretion *Proc Soc Exp Biol Med* 49 87-90, 1942

KIIL, F *The Function of the Ureter and Renal Pelvis* Philadelphia Saunders 1957

LAPIDES, J *Urology*, edited by M F Campbell and J H Harrison Philadelphia Saunders 1970 pp 1343-1378

LAPIDES, J E P AJEMIAN, B H STEWART, B A BREAKEY AND J R LICHTWARDT Further observations on the kinetics of the urethrovesical sphincter J Urol 84 86-94 1960

LASSITER, W E, C W GOTTSCHALK, AND M MYLLE Micro puncture study of net transtubular movement of water and urea in nondiuretic mammalian kidney *Amer J Physiol* 200 1139-1147, 1961

LEVINSKY, N G, AND R W BERLINER The role of urea in the urine concentrating mechanism *J Clin Invest* 38 741-748, 1959

LOTSPEICH W D Renal tubular reabsorption of inorganic sulfate in the normal dog *Amer J Physiol* 151 311-318 1947

LUDWIG C Nieren und Harnbereitung In *Handwörterbuch der Physiologie* edited by R Wagner Braunschweig Germany Fr Bieweg u Sohn 1844 vol 2 p 628-640

MALNIC G M DE MELLO AIRES AND G GIEBISCH Potassium transport across distal tubules of rat kidney during alterations of acid base equilibrium *Amer J Physiol* 221 1192-1208 1971

MARSHALL F K JR AND J L VICKERS The mechanism of the elimination of phenolsulphonphthalein by the kidney a proof of secretion by the convoluted tubules *Johns Hopkins Med J* 34 1-7 1923

MARTINO J A AND L E EARLEY Demonstration of a role of physical factors as determinants of the natriuretic response to volume expansion *J Clin Invest* 46 1963-1978 1967

MÖLLER E J F McINTOSH AND D D VAN SLYKE Studies of urea excretion II Relationship between urine volume and the rate of urea excretion by normal adults *J Clin Invest* 6 427-465 1929

MONTGOMERY H AND J A PIERCE The site of acidification of the urine within the renal tubule in Amphibia *Amer J Physiol* 118 144-152 1937

MORGAN T, AND R W BERLINER Permeability of the loop of Henle vasa recta and collecting duct to water urea and sodium *Amer J Physiol* 215 108-115 1968

MUELLER, C B A SURTSHIN, M R CARLIN AND H L WHITE Glomerular and tubular influences on sodium and water excretion *Amer J Physiol* 165 411-422 1951

NASH T P AND S R BENEDICT The ammonia content of the blood and its bearing on the mechanism of acid neutralization in the animal organism *J Biol Chem* 48 463-488, 1921

NELP W B H N WAGNER, JR AND R C REBA Renal excretion of vitamin B_{12} and its use in measurement of glomerular filtration rate in man *J Lab Clin Med* 63 480-491 1964

ORLOFF, J The role of the kidney in the regulation of acid base balance *Yale J Biol Med* 29 211-228, 1956

ORLOFF J AND R W BERLINER The mechanism of the excretion of ammonia in the dog *J Clin Invest* 35 223-235 1956

ORLOFF J T J KENNEDY, JR AND R W BERLINER. The effect of potassium in nephrectomized rats with hypokalemic alkalosis *J Clin Invest* 32 538-542, 1953

PAPPENHEIMER J R E M RENKIN AND L M BORRERO Filtration diffusion and molecular sieving through peripheral capillary membranes A contribution to the pore theory of capillary permeability *Amer J Physiol* 167 13-46, 1951

PITTS, R F A comparison of the renal reabsorptive processes for several amino acids *Amer J Physiol* 140 535-547 1944

PITTS R F Renal excretion of acid *Fed Proc* 7 418-426, 1948

PITTS R F Renal production and excretion of ammonia *Amer J Med* 36 720-742 1964

PITTS, R F, AND R S ALEXANDER The renal reabsorptive mechanism for inorganic phosphate in normal and acidotic dogs *Amer J Physiol* 142 648-622 1944

PITTS, R F, AND R S ALEXANDER. The nature of the renal tubular mechanism for acidifying the urine *Amer J Physiol* 144 239-254 1945

PITTS, R F J L AYER AND W A SCHIESS The renal regulation of acid-base balance in man III The reabsorption and excretion of bicarbonate *J Clin Invest* 28 35-44, 1949

PITTS, F R, AND W D LOTSPEICH Bicarbonate and the renal regulation of acid base balance *Amer J Physiol* 147 138-154, 1946

PITTS, R F, W D LOTSPEICH, W A SCHIESS, AND J L AYER The renal regulation of acid base balance in man I The nature of the mechanism for acidifying the urine *J Clin Invest* 27 48-56, 1948

RECTOR, F C, JR, AND J ORLOFF The effect of the administration of sodium bicarbonate and ammonium chloride on the excretion and production of ammonia The absence of alterations in the activity of renal ammonia producing enzymes in the dog *J Clin Invest* 38 366-372, 1959

REHBERG, P B Studies on kidney function I The rate of filtration and reabsorption in the human kidney *Biochem J* 20 447-460 1926

RELMAN, A S, B ETSTEN, AND W B SCHWARTZ The regulation of renal bicarbonate reabsorption by plasma carbon dioxide tension *J Clin Invest* 32 972-978, 1953

RHODIN J A G Structure of the kidney In *Diseases of the Kidney* EDITED BY M B Strauss and L G Welt Boston Little, Brown 1963, p 1-29

RICHARDS A N The Croonian Lecture Processes of urine formation *Proc Roy Soc (London)* [*Biol*] *Ser B*126 398-432 1938

RICHARDS A N P A BOTT, AND B B WESTFALL Experiments concerning the possibility that inulin is secreted by the renal tubules *Amer J Physiol* 123 281-298 1938

RICHARDS A N AND A M WALKER Methods of collecting fluid from known regions of the renal tubules of Amphibia and of perfusing the lumen of a single tubule *Amer J Physiol* 118 111-120, 1936

SARTORIUS O W J C ROEMMELT AND R F PITTS The renal regulation of acid base balance in man IV The nature of the renal compensations in ammonium chloride acidosis *J Clin Invest* 28 423-439, 1949

SAWYER W H Increased renal reabsorption of osmotically free water by the toad (*Bufo marinus*) in response to neurohypophysial hormones *Amer J Physiol* 189 564-568, 1957

SHANNON J A The excretion of inulin by the dog *Amer J Physiol* 112 405-413, 1935

SHANNON J A Glomerular filtration and urea excretion in relation to urine flow in the dog *Amer J Physiol* 117 206-225 1936

SHANNON J A Urea excretion in the normal dog during forced diuresis *Amer J Physiol* 122 782-787 1938

SHANNON, J A Renal tubular excretion *Physiol Rev* 19 63-93, 1939

SHANNON, J A, AND S FISHER The renal tubular reabsorption of glucose in the normal dog *Amer J Physiol* 122 765-774 1938

SHANNON, J A, AND H W SMITH The excretion of inulin, xylose and urea by normal and phlorizinized man *J Clin Invest* 14 393-401 1935

SMITH, H W *The Physiology of the Kidney* New York Oxford Univ Press, 1937

SMITH, H W *Lectures on the Kidney* Lawrence Kansas Univ Ext Div, Univ Kansas 1943

SMITH, H W *The Kidney Structure and Function in Health and Disease* New York Oxford Univ Press, 1951

SMITH, H W Renal excretion of sodium and water *Fed Proc* 11 701-705, 1952

SMITH H W, N FINKELSTEIN L ALIMINOSA, B CRAWFORD, AND M GRABER The renal clearances of substituted hippuric acid derivatives and other aromatic acids in dog and man *J Clin Invest* 24 388-404, 1945

SMITH, H W, W GOLDRING, AND H CHASIS The measurement of the tubular excretory mass, effective blood flow and filtration rate in the normal human kidney *J Clin Invest* 17 263-278, 1938

SMITH, H, W GOLDRING H CHASIS, H A RANGES AND S E BRADLEY The application of saturation methods to the study of glomerular and tubular function in the human kidney *J Mount Sinai Hosp N Y* 10 59-108 1943

SMITH W W N FINKELSTEIN AND H W SMITH Renal excretion of hexitols (sorbitol, mannitol and dulcitol) and their derivatives (sorbitan, isomannide and sorbide) and of endogenous creatinine-like chromogen in dog and man *J Biol Chem* 135 231-250 1940

SULLIVAN W J AND P J DORMAN The renal response to chronic respiratory acidosis *J Clin Invest* 34 268-276 1955

THATCHER, J S AND A W RADIKE Tolerance to potassium in toxication in the albino rat *Amer J Physiol* 151 138-146 1947

TRUETA J A E BARCLAY P DANIEL K J FRANKLIN AND M M L PRITCHARD Renal pathology in the light of recent neurovascular studies *Lancet* 2 237-238 1946

USSING H H The distinction by means of tracers between active transport and diffusion The transfer of iodide across the isolated frog skin *Acta Physiol Scand* 19 43-56, 1949

VAN SLYKE, D D R A PHILLIPS, P B HAMILTON, R M ARCHIBALD, P H FUTCHER AND A HILLER Glutamine as source material of urinary ammonia *J Biol Chem* 150 481-482 1943

VERNEY, E B Antidiuretic hormone and the factors which determine its release *Proc Roy Soc (London)* [*Biol*], *Ser B* 135 25-106 1947

WALKER, A M, P A BOTT J OLIVER, AND M C McDOWELL The collection and analysis of fluid from single nephrons of the mammalian kidney *Amer J Physiol* 134 580-595 1941

WALLENIUS, G Renal clearance of dextran as a measure of glomerular permeability *Acta Soc Med Upsal* 59 (Suppl 4) 1-91 1954

WEINER I M, AND G H MUDGE Renal tubular mechanisms for excretion of organic acids and bases *Amer J Med* 36 743-762 1964

WESSON, L G, JR AND W P ANSLOW JR Excretion of sodium and water during osmotic diuresis in the dog *Amer J Physiol* 153 465-474 1948

WINDHAGER, E E G WHITTEMBURY, D E OKEN, H J SCHATZMANN AND A K SOLOMON Single proximal tubules of the Necturus kidney III Dependence of H_2O movement on NaCl concentration *Amer J Physiol* 197 313-318, 1959

WIRZ, H Der osmotische Druck in den corticalen Tubuli der Rattenniere *Helv Physiol Pharmacol Acta* 14 353-362, 1956

WIRZ, H B HARGITAY, AND W KUHN Lokalisation des Konzentrierungsprozesses in der Niere durch direkte Kryoskopie *Helv Physiol Pharmacol Acta* 9 196-207 1951

WOODBURNE, R T Structure and function of the urinary bladder *J Urol* 84 79-84 1960

WOODBURNE, R T Anatomy of the ureterovesical junction *J Urol* 92 431-435 1964

6

Respiration

1 / UPTAKE AND DELIVERY OF THE RESPIRATORY GASES

The cells of higher animals continually use oxygen and produce carbon dioxide as a waste product Breathing movements promote the transfer of oxygen from the atmosphere into the lungs and the removal of carbon dioxide from the lungs Oxygen diffuses into the blood in the pulmonary capillaries to be carried to all parts of the body, where it diffuses from the capillary blood into the interstitial fluid and on to the cells Carbon dioxide diffuses from the cells through the interstitial fluid to the blood in the systemic capillaries, and then is carried to the lungs where it diffuses into the alveolar air The continual processes of respiration, regulated by a complex of "feed-back" and other mechanisms, play a major role in the maintenance of homeostasis Under constant conditions, the respiratory processes tend to attain a condition which may be called a "steady state," but not an "equilibrium with the environment," since at equilibrium all reactions would be at a standstill

The major problems to be considered under the heading of respiratory physiology are *1*) the uptake of oxygen from alveolar air and its delivery to the tissues, *2*) the uptake of carbon dioxide from tissues and its delivery to the alveolar air, *3*) the mechanism of rhythmic inspiration and expiration, and *4*) the mechanisms by means of which the amount of air breathed per minute is adjusted to the needs of the body The actual chemical processes by which oxygen is utilized in the cells (cellular respiration) will not be discussed, since this subject is considered in detail in biochemistry texts Some of the derangements of respiration are considered under the appropriate major headings, while others of a more general nature are described in the final chapter of this section

It is assumed that the reader already will have knowledge of elementary biochemistry and human anatomy Further, it is essential that the relation of carbon dioxide to the acid-base balance (sect 4, chap 11), the physiology of nerve, muscle, and reflex action (sect 9, chap 4) and the physiology of the autonomic nervous system (sect 9, chap 3) should be studied in advance in order to understand the chemical and neural control of respiration

A system of notation for use in respiratory physiology, summarized in Table 6 1, was agreed on by a group of physiologists meeting in Atlantic City in 1950 This system is now generally accepted, and will be followed in this section

THE KINETIC THEORY OF GASES

A quantity of gas when placed in a container of any size expands until it fills the confining vessel This fact is explained upon the theory that the gas molecules are in continuous motion, moving through space at high velocity and being deflected from their course only upon coming into collision with other gas molecules or with the boundaries of the space itself which they strike and from which they rebound repeatedly Such movements constitute a bombardment upon the confining walls which is responsible for the pressure of the gas The greater the number of molecules in any given space, and the higher their mean velocity, the more frequent will be the impacts

Boyle's Law If the volume of a given mass of gas is reduced, the molecules are brought closer together and the rate of bombardment upon a unit surface increases, the increase becoming manifest as a rise in pressure Robert Boyle first noted the quantitative relationship between pressure and volume *at a constant temperature the pressure of a given mass of gas is inversely proportional to its volume*, i e , the product of the pressure and the volume remains constant

Avogadro's Law states that *equal volumes of different gases at the same temperature and pressure contain the same number of molecules* Thus a gram-molecule of any gas at STPD occupies a volume of 22 4 liters Since the force of an impact is related to the mass of the molecule as well as to its velocity, it follows that at a given temperature the molecules of a gas of low molecular weight are traveling faster than are those of a heavier gas The mean kinetic energy of the gas molecules has a certain value which is related to temperature alone

Table 6 1[a]

Symbols and abbreviations used by pulmonary physiologists (based on report in Federation Proceedings 9 602-605 1950)

SPECIAL SYMBOLS

—Dash above any symbol indicates a *mean* value
Dot above any symbol indicates a *time derivative*

For Gases				For Blood			
Primary Symbols (large capital letters)		Examples		Primary Symbols (large capital letters)		Examples	
V	= gas volume	V_A	= volume of alveolar gas	Q	= volume of blood	Qc	= volume of blood in pulmonary capillaries
$\dot{V}$	= gas volume/unit time	$\dot{V}o_2$	= O_2 consumption/min	$\dot{Q}$	= volume flow of blood/unit time	$\dot{Q}c$	= blood flow through pulmonary capillaries/min
P	= gas pressure	$P_{A\,O}$	= alveolar O_2 pressure	C	= concentration of gas in blood phase	Ca_{O_2}	= ml O_2 in 100 ml arterial blood
$\bar{P}$	= mean gas pressure	$\bar{P}c_{\bar{Q}_2}$	= mean capillary O_2 pressure	S	= % saturation of Hb with O_2 or CO	Sv_{O_2}	= saturation of Hb with O_2 in mixed venous blood
F	= fractional concentration in dry gas phase	$F_{I_{O_2}}$	= fractional concentration of O_2 in inspired gas				
f	= respiratory frequency (breaths/unit time)			Secondary Symbols (small letters)		Examples	
D	= diffusing capacity	D_{O_2}	= diffusing capacity for O_2 (ml O_2/min/mm Hg)	a	= arterial blood	Pa_{CO_2}	= partial pressure of CO_2 in arterial blood
R	= respiratory exchange ratio	R	= Vco_2/Vo_2	v	= venous blood	$P\bar{v}_{O_2}$	= partial pressure of O_2 in mixed venous blood
Secondary Symbols (small capital letters)		Examples		c	= capillary blood	Pc_{CO}	= partial pressure of CO in pulmonary capillary blood
I	= inspired gas	$F_{I_{CO_2}}$	= fractional concentration of CO_2 in inspired gas	b	= blood in general		
E	= expired gas	V_E	= volume of expired gas	s	= steady state		
A	= alveolar gas	V_A	= alveolar ventilation/min				
T	= tidal gas	V_T	= tidal volume				
D	= dead space gas	V_D	= volume of dead space gas				
B	= barometric	P_B	= barometric pressure				
STPD	= 0°C, 760 mm Hg, dry						
BTPS	= body temperature and pressure saturated with water vapor						
ATPS	= ambient temperature and pressure saturated with water vapor						

[a] After Comroe et al (1962)

Law of Charles (or Gay-Lussac) A rise in temperature increases the velocity and kinetic energy of the gas molecules The force of the impacts is thus increased, and at constant volume the rate of bombardment also is increased For each rise in temperature 1° C a gas kept at constant volume increases its pressure by $\frac{1}{273}$ of its pressure at 0° C If, instead, the pressure is held constant, the volume increases by a corresponding amount Taking −273° C as the absolute zero, it follows that *the pressure of a gas at con-*

stant volume (or the volume at constant pressure) is proportional to the absolute temperature

The ideal gas law is a quantitative expression of the foregoing principles

$$PV = nRT$$

where P is the pressure, V is the volume, n is the number of gram-molecules of the gas, R is the "gas constant," and T is the absolute temperature When the units employed are atmospheres, liters and centigrade degrees, $R = 0.082$ For any real gas this expression is subject to small corrections related to the actual volume of the molecules themselves and the attractive forces between them

Diffusion In all circumstances the gas molecules, as a result of their movement, will distribute themselves evenly throughout the space in which they are confined, and consequently the pressure will become equal throughout all parts of the space This tendency of a gas to pass from a point of higher pressure to a point of lower pressure, and to distribute itself throughout the available space is called diffusion However, the term applies equally to the distribution of dissolved molecules or ions in solution, and to the transfer of molecules across gas-liquid interfaces The rate of diffusion of a *gas* at a given temperature is inversely proportional to the square root of its density Thus hydrogen diffuses four times as rapidly as does oxygen

Dalton's Law of Partial Pressures If we should deal with a mixture of two or more gases, it would be found that each component of the mixture behaved as though present alone Its molecules would become distributed evenly throughout the mixture, although its rate of diffusion would be retarded by collisions with other molecules *Each gas in a mixture exerts a pressure according to its own concentration, independently of the other gases present* The pressure of each constituent is referred to as its *partial pressure* or *tension* and the total pressure is the *sum of the partial pressures* of all the gases present

Henry's Law When a gas is brought into contact with water, gas molecules bombard the surface of the liquid and pass into it, but the dissolved gas molecules also tend to escape from the liquid After a time their rate of escape equals their rate of entry When this equilibrium is reached, the solution is said to be saturated Each gas has its own characteristic solubility at a given temperature and pressure If no chemical reaction is involved, but only physical solution, Henry's Law states that *the concentration of a gas in solution at equilibrium is directly proportional to its partial pressure in the gas phase* The solubility of oxygen, for example, is described by the expression

$$Co_2 = \alpha O_2 \, Po_2$$

where Co_2 is its concentration in aqueous solution, Po_2 is its tension, and αO_2 is a constant which is characteristic of oxygen If Co_2 is expressed in milliliters (STPD) per milliliter of solution and Po_2 in atmospheres, αO_2 represents the solubility in milliliters of oxygen gas per ml of water and per atmosphere This is called the solubility coefficient or absorption coefficient of oxygen

The solubility of a gas is not affected by the presence of other gases, each gas in a mixture dissolves in accordance with its own partial pressure and absorption coefficient The solubility is decreased slightly, however, by the presence of salts or other solutes Hence the absorption coefficients are slightly lower for solution in blood plasma than for solution in water (Table 6 2) The values for whole blood are influenced also by the presence of the erythrocytes In general, the solubility of a gas is decreased by an increase in temperature

Although carbon dioxide in solution combines with water to form H_2CO_3, this reaction does not cause deviations from Henry's Law The two forms are present in a constant ratio, and the equilibrium is such that only one part in 680 is in the hydrated form at 38° C (Roughton, 1943-1944)

When the absorption coefficient is known, it is a simple matter to calculate the tension from the milliliters of gas per milliliters of solution and vice versa Hence it is often convenient to use the value representing tension (in atmospheres or, more commonly, in mm Hg) as a measure of the concentration of physically dissolved gas in a solution If the liquid after equilibration were separated from the gas phase, this would not change the tension of the dissolved gas, hence the use of this measure does not require the presence of a gas phase If the solution were subsequently exposed to a vacuum or to an atmosphere of other gases, however, the dissolved gas would tend to diffuse out of the solution Hence the tension of a dissolved gas can be regarded as a measure of its tendency to escape from solution

Aqueous tension Water itself, when exposed to a vacuum or to an atmosphere of other gases, tends to evaporate until an equilibrium is reached between the gas phase and the liquid Although it is common to speak of the air becoming saturated, the tension or vapor pressure of water attained at equilibrium is not

Table 6 2

Absorption coefficients of gases in water blood plasma and whole blood in milliliters per milliliters and per atmosphere

	αO_2	αCO_2	αN_2
Water, 20° C	0 0314[a]	0 872[a]	0 0155[a]
Water, 37° C	0 0237[b]	0 547[b]	0 0130[c]
Plasma, 37° C	0 0213[b]	0 521[b]	0 0119[c]
Blood, 37° C	0 0235[b]	0 48[b]	0 0132[c]

[a] Radford, 1964

[b] Rossier et al , 1960

[c] Van Slyke et al , 1934 (corrected to 37° C)

affected by the presence or absence of other gases It depends simply upon the temperature At body temperature (37° C) the aqueous tension of water is 47 mm Hg Substances in solution reduce the aqueous tension by small amounts

The Exchange of Gases in the Lungs

CONCEPT OF TENSION GRADIENTS

Oxygen from the air in the alveoli enters the blood in the pulmonary capillaries by diffusion through the alveolar and capillary membranes, while carbon dioxide diffuses in the reverse direction The transfer of a quantity of gas by diffusion can occur only from a place of higher to a place of lower concentration, i e , in the direction of a *concentration gradient* Strictly speaking, such a gradient would be defined as a decrease in concentration per unit of distance However, the actual distance traversed may not be precisely known and may not be of major interest It is customary to refer to the total difference in oxygen tension between alveolar air and pulmonary blood plasma as the oxygen diffusion gradient The direction of this gradient is from the higher (alveolar) toward the lower (capillary) tension The carbon dioxide gradient, stated in the same terms, is in the opposite direction Similar gradients exist between the blood plasma and the interstitial fluid, and in turn between the interstitial fluid and the interior of the respiring cells In each case, the tension gradients are to be regarded as the driving forces for the exchange of gases Combined forms of oxygen and carbon dioxide existing in the blood play no direct role in these exchanges, they serve only as reservoirs which, by dissociation and the laws of mass action, replenish and maintain the tensions of the physically dissolved gases

The rate of diffusion of a gas through a liquid per unit of *tension* gradient is directly proportional to the solubility of the gas and inversely proportional to the square root of its molecular weight Hence, at a given tension gradient and at body temperature, the diffusion rate of carbon dioxide in solution is about 20 times that of oxygen

THE ALVEOLAR AIR

The mixture of gases in the alveoli consists of inhaled air from which oxygen is being removed and to which carbon dioxide and water vapor are being added Under resting conditions each inhalation brings into the respiratory system a volume of air representing less than 10% of the total lung capacity, and not all of this reaches the alveoli Hence the composition of the alveolar air shows only minor fluctuations with the respiratory cycle The composition is not exactly the same in all parts of the lung, since there are variations in the effectiveness of ventilation The average composition, however, is maintained approximately constant by the mechanisms regulating the respiratory rate and volume

A sample of expired air taken near the end of a normal expiration (end-tidal sample) is fairly representative of the average composition of alveolar air The *Rahn sampler* is a device which automatically collects end-tidal samples, permitting continuous analysis and recording of the alveolar air (Rahn and Otis, 1949) In the older technique of *Haldane and Priestley*, allowance is made for small differences occurring during the respiratory cycle by taking two samples one from the last part of a forced expiration following a normal inspiration, the other from a forced expiration following a normal expiration The average of the two is taken to represent the mean composition of the alveolar air (Rossier et al , 1960) The *"functional" alveolar oxygen tension* can be obtained by an indirect method based on the *"alveolar air equation"*, this calculation requires analysis of the expired air (collected over a period of about 5 min) and measurement of the arterial CO_2 tension It is assumed that the mean CO_2 tension of the alveolar air is equal to that of the arterial blood (Riley et al , 1946, Otis, 1964)

The air in the lungs has a temperature of about 37° C and is usually considered to be saturated with water vapor Its aqueous tension is therefore 47 mm Hg [1] When the air cools after exhalation, some of the water vapor condenses The tension of water vapor in room air at 22° C could be no higher than 20 mm Hg and is usually much lower

In the measurement of respiratory gases, the volumes are expressed dry (i e , less the aqueous vapor) and at standard temperature and pressure The individual gases are ex-

[1] Christie and Loomis (1932), from direct measurements, obtained a slightly lower value for the aqueous tension of alveolar air (45 mm Hg) They claim that the alveolar air is not fully saturated and that the temperature of the lung is lower than is usually assumed Hyperpnea reduced the aqueous tension by as much as 7 mm Hg Holding the breath increased it by 0 5 mm Hg

pressed as percentages of the dry volume In order to arrive at the *tension* of a gas from its percentage by analysis, the figure for the barometric pressure less the aqueous tension must be used in the calculation For example, if a sample of alveolar air (dry) is 5 6% CO_2, and the barometric pressure is 760 mm Hg, the *alveolar carbon dioxide tension* is calculated as follows

$$P_{ACO_2} = (5\ 6/100)(760 - 47) = 40 \text{ mm Hg}$$

In Table 6 3 the *average composition of alveolar air (at sea level) is compared with the known composition of dry atmospheric air* The latter does not show significant variations within the troposphere Values for oxygen, carbon dioxide and nitrogen tensions also are given

During voluntary hyperventilation the alveolar tension of CO_2 falls and that of oxygen rises When the breath is held or during periods of apnea, changes occur in the reverse direction At higher altitudes, the alveolar oxygen and nitrogen tensions are reduced along with the barometric pressure, but the alveolar P_{CO_2} does not show a proportionate decrease since it is controlled by the rate of ventilation of the lungs The *percentage* of CO_2 therefore increases as the barometric pressure decreases and the oxygen and nitrogen *percentages* must accordingly decrease When the P_{O_2} is reduced enough to reflexly increase the breathing, the P_{CO_2} begins to fall and the decrease in P_{O_2} is partially compensated

Although a small but constant amount of nitrogen is physically dissolved in the blood (in accordance with Henry's Law), gaseous nitrogen is neither used nor produced in metabolism Argon and other inert gases are usually included under "nitrogen," since no distinction need be made in discussions of respiratory physiology It will be noted that the percentage of nitrogen is slightly higher in alveolar air than in the inspired air This is due, not to any absolute increase in the quantity of nitrogen, but to the reduction of the total volume of the respiratory gases resulting from the fact that the oxygen consumption exceeds the CO_2 production

Table 6 3
Atmospheric and alvolar air

	O_2	CO_2	N_2[a]
Atmospheric air			
Per cent (dry)	20 95	0 04	79 01
Tension, mm Hg[b,c]	157	0 3	593
Alveolar air			
Per cent (dry)	14 0	5 6	80 4
Tension, mm Hg[b]	100	40	573

[a] Includes other inert gases, 0 93%
[b] At sea level, barometric pressure 760 mm Hg
[c] Assuming absolute humidity of 10 mm Hg

BLOOD GASES

Oxygen and Carbon Dioxide Tensions The mixed venous blood, which enters the pulmonary arteries and is carried to the lungs, has an oxygen tension of about 40 mm Hg The gradient in oxygen tension from alveolar air to the blood as it enters the pulmonary capillaries is therefore about 60 mm Hg With this gradient as the driving force, oxygen diffuses across the alveolar and capillary membranes Equilibrium is quickly established, and as the blood leaves the pulmonary capillaries it is nearly saturated with oxygen A sample of blood obtained from a systemic artery is found to have an oxygen tension slightly lower than that of the alveolar air, not because of failure to attain equilibrium across the pulmonary membranes, but for reasons to be discussed later

The venous blood has a CO_2 tension of about 46 mm Hg, hence at the entrance of a pulmonary capillary the diffusion gradient for the transfer of CO_2 to the alveolar air is only about 6 mm Hg However, since CO_2 is more soluble than oxygen, this gradient suffices for the removal from the blood of an amount of CO_2 almost equivalent to the oxygen absorbed, and equilibrium is reached before the blood leaves the pulmonary capillary Within the limit of error involved in alveolar air sampling, the CO_2 tension of arterial blood is found to equal that of the alveolar air [2]

The great rapidity with which these exchanges occur is evident from estimates of the time spent by an individual corpuscle in traversing a pulmonary capillary Roughton (1945) has estimated that the "transit time" is about 0 75 sec under resting conditions and is reduced to about half this value by strenuous exercise

[2] The end-expiratory CO_2 tension is measurably lower than the arterial value in shallow breathing and in pathological conditions characterized by significant alveolar dead space

Oxygen Content and Capacity The amounts of gases in blood are commonly expressed in terms of *volumes percent*, i e , milliliters of gas (STPD) per 100 milliliters of blood From the absorption coefficient of oxygen and the oxygen tension, it can be calculated that the physically dissolved oxygen in arterial blood amounts to only 0 3 volumes per cent If the transport of oxygen by the blood were dependent upon physical solution alone, an enormous cardiac output would be required to supply the amount of oxygen needed by the tissues Actually most of the oxygen is carried in combination with hemoglobin,[3] and the total *oxygen content* of the arterial blood is about 20 volumes per cent All of this oxygen can be removed from the blood by evacuation

The oxygen content of blood is dependent upon the oxygen tension, but the relationship between these two variables is not linear It is described by the *oxygen dissociation curve*, which will be discussed in detail in a later section If blood is equilibrated with air at 760 mm Hg, the hemoglobin becomes saturated The oxygen content of the blood, corrected by subtracting the small amount in physical solution, then represents the *oxygen capacity of the hemoglobin* The oxygen capacity is a measure of the active hemoglobin content, since 1 g of hemoglobin will combine with exactly 1 34 ml of oxygen (STPD) Hemoglobin values calculated from precise measurements of oxygen capacity are often used as a basis for standardizing colorimetric methods for the determination of hemoglobin, but it must be remembered that many blood samples contain traces of inactive hemoglobin compounds such as methemoglobin and carboxyhemoglobin

The ratio of the oxygen content (after correction for dissolved oxygen) to the oxygen capacity of the hemoglobin is a measure of the degree of saturation of the hemoglobin This ratio is usually multiplied by 100 to give the *per cent saturation* In arterial blood the hemoglobin is normally about 97% saturated The amount of oxygen given up by the blood as it passes through the tissues varies from one region to another and from time to time, depending upon the blood flow and the metabolic rate in the region concerned Under resting conditions the average arteriovenous difference is about 5 volumes per cent The mixed venous blood, then, has an oxygen content of about 15 volumes per cent and its hemoglobin is about 75% saturated Samples of mixed venous blood can be obtained from human subjects by cardiac catheterization

Average normal values for blood gases and hemoglobin content are summarized in Table 6 4

Carbon Dioxide Content The tension of CO_2 in the alveolar air is lower than that of oxygen, but the absorption coefficient of CO_2 is much the higher of the two The calculated amount of CO_2 present in *simple solution* in the arterial blood is 2 5 volumes per cent This is distributed between the red cells and the plasma

The total *carbon dioxide content* is the quantity of CO_2 that can be extracted from 100 ml of blood by exposure to a vacuum after the addition of acid In arterial blood this amounts to about 46 volumes per cent A large fraction of it exists in the form of *bicarbonate ions*, some of this is in the plasma and a smaller portion is inside the erythrocytes Another fraction is bound to hemoglobin in the erythrocytes as *carbaminohemoglobin*

These various forms of carbon dioxide exist in equilibrium with each other in the blood, and all are involved in the transport of CO_2 When CO_2 enters or leaves the blood, a rapid readjustment of the equilibrium occurs

Carbon Dioxide-Bicarbonate-pH Relationships The dissociation of a weak acid is described quantitatively by the *Henderson-Hasselbalch equation*, which is the logarithmic form of the law of mass action (see sect 4,

Table 6 4
Average normal analytical values for arterial and venous blood

	Arterial	Venous
O_2 tension, mm Hg	93	40
CO_2 tension, mm Hg	40	46
Dissolved O_2, vol %	0 3	0 1
Total O_2 content, vol %	19 8	15 2
O_2 capacity of Hb, vol %	20 1	20 1
O_2 saturation of Hb, %	97	75
Hemoglobin, g per 100 ml	15	15
Dissolved CO_2, vol %	2 5	2 9
CO_2 content, vol %	46	50
Plasma CO_2 content, vol %	56	60
Plasma CO_2 content, mM	25	27
Plasma pH	7 38	7 35

[3] See section 4, Chapter 4 for a discussion of the chemistry of hemoglobin

chap 11) As applied to the carbonic acid-bicarbonate system, this equation is

$$pH = pK + \log[HCO_3^-]/[H_2CO_3]$$

where pK is the negative logarithm of the dissociation constant of carbonic acid

Carbon dioxide forms carbonic acid by hydration

$$CO_2 + H_2O \rightleftharpoons H_2CO_3$$

The hydrated and unhydrated forms reach an equilibrium, and the dissolved carbon dioxide is the sum of the two forms However, the equilibrium point is so far to the left that only about 0 15% is in the hydrated form at 38° C Hence the concentration of the unhydrated form, designated $[CO_2]$, can be considered equal to the dissolved carbon dioxide with only a negligible error The small concentration of H_2CO_3 is directly proportional to $[CO_2]$ Hence if an appropriate value for K is used, the equation can be written

$$pH = pK' + \log[HCO_3^-]/[CO_2]$$

The established value for pK′ in *blood plasma* is 6 10[4] at 37° C When the plasma pH is 7 4, therefore, the ratio of plasma bicarbonate to dissolved carbon dioxide is 20 1 The concentrations of CO_2 and HCO_3^- can be expressed either in millimoles per liter or in volumes per cent of CO_2, the value for pK′ is 6 10 in either case

The dissolved carbon dioxide can be calculated from the CO_2 tension and the absorption coefficient α as determined on acidified plasma (Table 6 2) The plasma $[HCO_3^-]$ can be obtained by subtracting the dissolved CO_2 from the total measured CO_2 content of the plasma Incorporating these considerations, the equation becomes

$$pH = 6\,10 + \log([CO_2 \text{ content}] - \alpha P_{CO_2})/\alpha P_{CO_2}$$

If concentrations are expressed in volumes per cent and P_{CO_2} in millimeters of mercury, the value for α becomes 0 0686 In terms of millimoles per liter and mm Hg, α is 0 0308 (since 1 mole of CO_2 at STPD occupies 22 26 liters)

This equation is very useful since if any two of the three variables are measured the third can be calculated Its application to whole blood, however, is complicated since the carbon dioxide content of blood includes a considerable portion present as carbaminohemoglobin and the intracellular and extracellular concentrations of H^+ and HCO_3^- are complexly related

[4] Calculated from precise measurements of P_{CO_2}, pH, and CO_2 content or bicarbonate in plasma K′ therefore is a composite constant including the true dissociation constant of H_2CO_3, the equilibrium constant for the hydration of CO_2, the activity coefficient of the bicarbonate ion, and a small correction for equating $[CO_2]$ with the dissolved carbon dioxide The pK′ value varies slightly with pH, probably because traces of carbonate and of plasma carbaminobound carbon dioxide are included in the measurements of bicarbonate (Severinghaus et al , 1956, Siggard Anderson, 1962a)

Measuring Respiratory Gases

For many years the most widely used methods for the measurement of respiratory gases were those developed by J S Haldane at Oxford and by D D Van Slyke and his associates at The Rockefeller Institute Numerous modifications of these methods have been described, and micro methods based on the sample principles have been introduced These methods depend upon the careful measurement of gas volumes and pressures and the selective absorption of carbon dioxide and of oxygen by appropriate reagents In the analysis of blood the gases are first released by chemical treatment and vacuum extraction

More recently, various electronic and other instruments have been devised for the measurement or continuous monitoring of oxygen, CO_2, and nitrogen in gas mixtures, and of oxygen tension, hemoglobin saturation, and CO_2 tension in blood Some of the more important methods now in use will be outlined briefly Detailed descriptions of these are available in a compilation by Consolazio et al (1963) or in other references cited

ANALYSIS OF GAS MIXTURES

Haldane Method The gas sample is held in a sampling tube over mercury or in an oiled syringe A 10-ml volume is drawn into the gas buret and measured at the prevailing atmospheric pressure and temperature CO_2 is absorbed by strong potassium hydroxide solution, and oxygen by an alkaline solution of potassium pyrogallate or of sodium hydrosulfite and sodium anthraquinone-β-sulfonate, the respective decreases in volume are recorded The percentage of nitrogen is calculated from the volume of gas remaining unabsorbed The inside of the buret is kept wet with dilute sulfuric acid to maintain a constant aqueous tension throughout, the results then represent percentages on a dry basis

Scholander Method This is a rapid micro adaptation of the Haldane method employing an apparatus of very different design A 0 5-ml gas sample is introduced into a small reaction chamber connected to a micrometer buret The buret is filled with mercury The pressure of the sample is balanced against that of air in a compensating chamber by means of an indicating drop of fluid in a capillary tube connecting the two chambers Absorbing solutions for CO_2 and oxygen can be tilted into the reaction chamber without altering the total volume While a gas is being absorbed, mercury is forced into the reaction chamber from the micrometer buret at a rate sufficient to maintain a constant pressure The volume of mercury introduced is then measured in terms of micrometer divisions

The Pauling Oxygen Analyzer This instrument measures the partial pressure of oxygen in a gas mixture Its operation depends on the fact that oxygen is paramagnetic while most other gases are diamagnetic, i e , the magnetic susceptibility of oxygen is very much greater than that of any other common gas A small glass dumbbell filled with nitrogen or oxygen (depending on the range of the instrument) is suspended between the poles of a magnet by attachment to a vertically stretched silica fiber By rotating around the fiber, the dumbbell can move in the magnetic field The incoming gas surrounds the dumbbell, and the oxygen is attracted to the strongest part of the magnetic field The dumbbell is subjected to a magnetic force depending on the difference between the magnetic susceptibility of its spheres and that of the gas which the spheres displace, and rotates as the spheres move farther into or out of the magnetic field An equilibrium position is reached at which the magnetic rotational force is balanced by the restoring force of the twisted fiber Thus the operation of the instrument is analogous to the measurement of the density of a liquid by weighing a solid body suspended in the liquid A small mirror attached to the dumbbell reflects a light beam onto a scale which is calibrated to indicate oxygen tension The Pauling analyzer has proved to be particularly useful for the analysis of expired air in studies of metabolism employing the open-circuit method

The Infrared CO_2 Analyzer CO_2 absorbs infrared radiation in proportion to its concentration Rapidly responding instruments based on this property can be used for continuous monitoring of carbon dioxide in the expired air. The gas mixture is exhaled through a mouthpiece connected to the sampling tube, or a continuous sample is drawn into the analyzer through a very small polyethylene catheter A beam of infrared radiation passes through the sample and then into a detector connected to an amplifier-recorder system The instrument registers the infrared absorption and can be calibrated in terms of per cent CO_2 (Collier et al , 1955)

The Nitrogen Analyzer Gases subjected to electrical excitation ionize and emit radiation The nitrogen analyzer contains an electrical discharge tube through which a slow, steady stream of the gas mixture is drawn at very low pressure A high potential is applied and the radiation at wavelengths characteristic of nitrogen is passed by a suitable filter and allowed to activate a photoelectric cell The signal is amplified and indicated on a meter or recorder calibrated in per cent nitrogen The response of the instrument is extremely rapid

Gas Chromatography This recently developed technique seems destined to largely supersede other methods for the analysis of individual gas samples The apparatus for respiratory gases employs gas-solid chromatography, and usually includes two columns connected in series or in parallel, each with its own thermal conductivity detector The gas sample is injected into a stream of helium entering the columns CO_2 is separated from the other respiratory gases in one column, and, as the CO_2-laden portion of the helium carrier passes through the detector, a variation in thermal conductivity is recorded as a peak, the size of which is proportional to the percentage of CO_2 in the injected sample Oxygen and nitrogen are separated and measured in the other column (Lysyj and Newton, 1963)

Other Methods The Van Slyke apparatus, the polarographic method for oxygen and the CO_2 electrode are applicable to gas mixtures, but more commonly are applied to the analysis of blood or plasma These methods are described in the next section The mass spectrograph also is sometimes used for the analysis of respiratory gases

GASES IN BLOOD OR PLASMA

Handling Blood Samples Blood for gasometric studies must be drawn and handled without exposure to air It is sometimes drawn under oil or over mercury, but more commonly into a syringe which is immediately closed by replacing the needle with a cap The syringe may then be stored briefly in ice water prior to the analysis The sampling tube or syringe must contain an anticoagulant, usually a drop of heparin solution When plasma is required, it is obtained by centrifuging the blood under a layer of mineral oil or, preferably, in a closed syringe with the plunger held in position by a plastic case fitted around its exposed portion (Peters and Van Slyke, 1932, Comroe, 1950)

Tonometers As originally defined, a tonometer is an instrument for measuring tension However, the first methods used for measuring gas tensions in blood required equilibration of the blood with a gas phase and analysis of the gas Hence the vessel in which the equilibration was brought about came to be called a tonometer Many forms of apparatus have

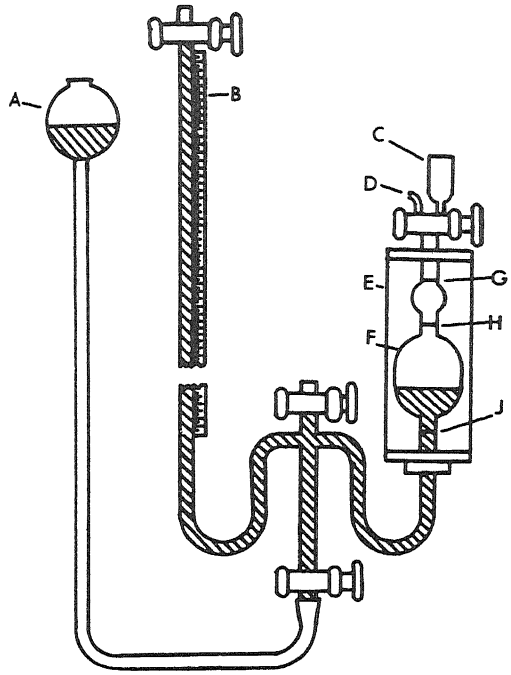


Fig 6 1 Van Slyke-Neill manometric apparatus *A* leveling bulb *B* manometer *C* cup *D* waste tube *E* water jacket *F* reaction chamber *G* 0 5 ml mark *H* 2 0 ml mark *J* 50 ml mark

been devised for this equilibration For example, in experiments to establish the oxygen dissociation curve it is desirable to expose small volumes of blood to large volumes of gas mixtures This can be done in a closed space, or a stream of gas can be passed over the blood On the other hand, in the "bubble" method for measuring gas tensions a relatively large volume of blood is exposed to a bubble of air so small that the exchange of gases has a negligible effect on the composition of the blood

The Van Slyke-Neill Manometric Apparatus (Fig 6 1) The analysis of blood or plasma is carried out by filling the reaction chamber with mercury and drawing in a 0 1 to 2 0-ml sample, preceded or followed by reagents to free the gases Carbon dioxide is released by lactic acid, and oxygen by potassium ferricyanide with saponin added to induce rapid hemolysis The gases are extracted by lowering the mercury to the 50-ml mark to create a vacuum and agitating the contents of the chamber with a shaking motor or a magnetic stirrer The volume of the extracted gases is then reduced to 0 5 ml or 2 0 ml and the corresponding pressure is read on the manometer A small amount of sodium hydroxide is introduced into the chamber to absorb the CO_2, and from the observed decrease in pressure at constant volume the CO_2 *content* of the sample is calculated The *oxygen content* is determined in the same way by absorption with a sodium hydrosulfite-sodium anthraquinone-β-sulfonate mixture

By this method the oxygen and carbon dioxide contents can be measured on a single blood sample *Oxygen capacity* is determined on a separate sample by first equilibrating the blood with air to saturate the hemoglobin and then measuring its oxygen content The *per cent saturation of hemoglobin* in the original sample can then be calculated The *oxygen tension* can be determined from the standard oxygen dissociation curve, provided the pH of the blood also is carefully measured and the necessary correction applied (Severinghaus, 1958) This method gives reliable values for oxygen tension in the range between 0 and about 60 to 70 mm Hg, but at higher levels the curve becomes too flat to allow accurate readings

The *carbon dioxide tension* and the *plasma bicarbonate*[5] can be calculated from the carbon dioxide content and the pH of the plasma by use of the Henderson-Hasselbalch equation Either plasma or whole blood may be used, but loss of CO_2 must be avoided The *plasma CO_2 content* may be measured on separated plasma or may be calculated from the analysis of whole blood However, if whole blood is used it is necessary to determine the hematocrit (or the oxygen capacity) and the per cent saturation of hemoglobin as well as the CO_2 content These data are needed since the CO_2 contained in the erythrocytes is proportional to the volume they occupy, and is influenced by the state of oxygenation of the hemoglobin as explained in a later section In practice the calculations are greatly simplified by the use of line charts, called nomograms, which graphically represent the Henderson-Hasselbalch equation and the relations between plasma and whole blood (Van Slyke and Sendroy, 1928, Singer and Hastings, 1948)

The Van Slyke volumetric apparatus operates on the same general principle as the manometric apparatus It is less accurate and less versatile than the

[5] The plasma bicarbonate represents the *alkali reserve* Van Slyke and Cullen introduced a determination of plasma carbon dioxide capacity or combining power which, although inexact, is often used clinically as a measure of the alkali reserve In this method the blood is drawn and centrifuged without strict anaerobic precautions, and the plasma is subsequently equilibrated with alveolar air at room temperature and analyzed for carbon dioxide content From this the bicarbonate concentration at 20° C is calculated The variable losses, being accompanied by shifts of bicarbonate into the cells, are not accurately replaced by the equilibration, but the method usually gives values somewhat higher than the original bicarbonate concentration since the solubility of carbon dioxide is higher at the lower temperature

manometric, but is simpler to operate, and is widely used for measuring blood gases in clinical laboratories After extraction, the gases are measured by observing the volume occupied at a fixed pressure (atmospheric), rather than the pressure exerted at a fixed volume (Peters and Van Slyke, 1932)

Ultramicro Gasometric Methods Natelson (1957) devised a manometric apparatus similar in principle to that of Van Slyke and Neill, but adapted for measurements on 0 03-ml quantities of blood or plasma More recently, Van Slyke and Plazin (1961) described a modification of the Van Slyke-Neill apparatus applicable to 0 025 to 0 030-ml samples Volumetric procedures on the same scale, developed by Roughton and Scholander, employ a device consisting of a capillary buret sealed to the nozzle of a 1-ml tuberculin syringe

Blood Gases by Gas Chromatography For the analysis of blood by this method it is only necessary to supplement the gas chromatograph (described previously) with a blood gas releasing system (Lenfant and Aucutt, 1966) The sample is mixed with releasing agents in a small chamber previously flushed with helium, and the evolved gases are carried into the chromatographic column by a stream of helium An analysis can be completed in 10 min or less

The "Bubble" Method for Oxygen and Carbon Dioxide Tensions A small bubble of air is equilibrated with blood at body temperature The bubble is then analyzed by micro methods, and from the percentages of oxygen and CO_2 the corresponding tensions are calculated If the volume of the bubble is kept very small in relation to the volume of blood, the amounts of oxygen and CO_2 exchanged during equilibration are insignificant and the results represent the original tensions in the blood This method was introduced by Krogh in 1908, and several appropriate forms of apparatus have been devised The Roughton-Scholander syringe analyzer is advantageous since it can be used for both the equilibration and the gas analysis (the "Riley technique")

Polarographic Measurement of Oxygen Tension If a platinum electrode in contact with a solution is made 0 6 to 0 7 volt negative to a chlorided silver electrode, dissolved oxygen molecules at the platinum surface receive electrons and are reduced to either H_2O_2 or OH^- The current passing from the platinum is linearly related to the availability of dissolved oxygen molecules at its surface When the platinum is in direct contact with the blood, however, the current flow is influenced by stirring and other factors which make adequate calibration difficult In the Clark electrode the electrical cell is separated from the blood by a thin polyethylene membrane that is permeable to oxygen (and CO_2) but impermeable to electrolytes and most other solutes If the platinum surface area is very small and the membrane is of the right thickness, this electrode shows a fairly rapid response, is insensitive to stirring, and can be calibrated to record *oxygen tension* in millimeters of mercury It can be used to analyze either solutions or gas mixtures, and for monitoring in vivo (Torres, 1963, Charlton et al , 1963)

The oxygen electrode has been applied to the measurement of *oxygen content* of blood by first mixing the sample with a ferricyanide reagent to release the oxygen into a known volume of solution The oxygen content is calculated from the tension observed and the solubility coefficient (Tucker, 1967, Cole and Hawkins, 1967)

The Carbon Dioxide Electrode This consists of a glass electrode, sensitive to pH changes, in contact with a thin layer of a solution containing a small concentration of bicarbonate ion and separated from the blood by a Teflon membrane Since the membrane is permeable to carbon dioxide but not to ions, the pH of the sample cannot affect the pH at the glass electrode surface The measured pH of the bicarbonate solution varies with the negative logarithm of the CO_2 *tension*, in accordance with the Henderson-Hasselbalch equation The electrode is calibrated against solutions of known CO_2 tension (Severinghaus and Bradley, 1958)

Interpolation Method for Carbon Dioxide Tension and Related Variables A graph showing the relation between log P_{CO_2} and the pH of a blood sample is a straight line By equilibrating a blood sample at two known P_{CO_2} levels and measuring the corresponding pH values, two points of the graph are determined and the curve for the sample can be drawn If the pH of the sample before equilibration is also measured, its P_{CO_2} can be found from the graph Carbon dioxide-oxygen mixtures are used for the equilibrations so that the hemoglobin will be fully oxygenated, a small estimated correction can be applied if the sample initially is not completely oxygenated Astrup (1961) described a capillary glass electrode and an equilibration unit, both thermostated at 38° C, with which the procedures can be carried out rapidly on a microscale

In addition to CO_2 tension, a number of other variables concerned with the acid-base balance can be calculated with the aid of a nomogram (Siggaard Anderson and Engel, 1960, Siggaard Anderson, 1962b) These include the standard bicarbonate of the plasma,[6] the buffer base and base excess (or deficit) of the blood, and the plasma bicarbonate, CO_2 content and CO_2 combining power The three pH values are the only measurements necessary

Photoelectric Oximetry An oximeter is a photoelectric photometer for measuring the fraction of he-

[6] Standard bicarbonate is defined as the bicarbonate concentration in the plasma of blood brought to a P_{CO_2} of 40 mm Hg at 38° C with the hemoglobin fully oxygenated This variable is useful in studies of nonrespiratory acid-base disorders since it is not influenced by variations in P_{CO_2} or oxygenation

moglobin in the oxygenated form Oxyhemoglobin absorbs less visible red light (620 to 770 mμ), and therefore transmits and reflects more in this waveband, than does reduced hemoglobin Infrared light of 805 mμ, however, is absorbed equally by oxygenated and reduced hemoglobin A transmission oximeter measures light transmitted by blood in both red and infrared regions, and the oxygen saturation is determined from the ratio of the two readings Some oximeters employ cuvettes in which the blood is contained or through which it is passed during either withdrawal from the body or diversion through an extracorporeal circuit Reflection oximeters also are available Instruments currently in use have been compared and evaluated by Cole and Hawkins (1967)

The ear oximeter is designed to record the oxygenation of the blood passing through the capillaries of the pinna The ear is warmed to some extent by the earpiece, with the result that a localized vasodilation occurs The capillary blood then is essentially arterial, since the oxygen utilization in this region is very small A transparent pneumatic pressure capsule is interposed between the light source and the ear so that allowance can be made for the light absorbed by the bloodless (pressurized) ear (Wood et al , 1960)

The Oxygen Dissociation Curve

The equilibrium between oxygen and hemoglobin is described by the oxygen dissociation curve, in which the per cent saturation of hemoglobin is plotted against the oxygen tension Since oxygen tension must be regarded as the *independent* variable, it is represented on the abscissa while per cent saturation (the *dependent* variable) is on the ordinate The equilibrium is influenced by electrolytes and other variables, hence curves determined on pure hemoglobin solutions differ in their exact shape and position from those determined on whole blood By equilibrating blood with known gas mixtures, Barcroft and Poulton (1913) determined the curves shown in Fig 6 2 For each curve the carbon dioxide was held constant while the oxygen tension was varied It is evident that increasing the CO_2 tension increases the dissociation, thus shifting the curve to the *right* (the "Bohr effect") This is due mainly to hydrogen ions produced by the dissociation of carbonic acid, but in part also to the formation of carbaminohemoglobin (Roughton, 1964) The hydrogen ion effect is related to the behavior of oxyhemoglobin and reduced hemoglobin as acids and the fact that the oxygenated form is a stronger acid than the reduced form, as discussed in a later section

An increase in temperature also shifts the curve to the right (increasing dissociation) Variations in blood electrolyte concentrations ordinarily are too small to have measurable effects on the oxygen dissociation curve

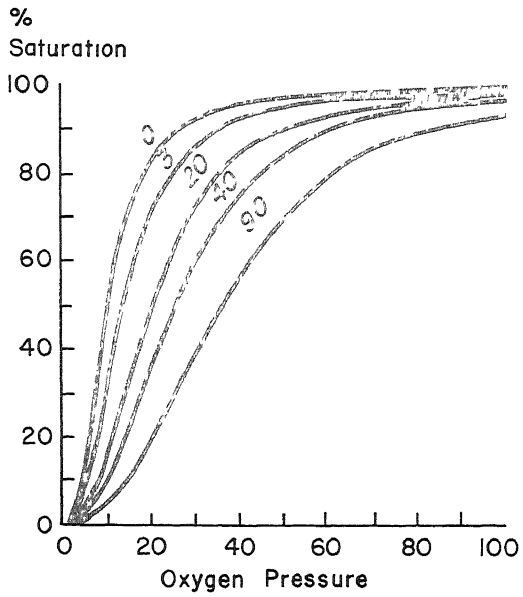


Fig 6 2 Dissociation curves of human blood exposed to 0 3 20 40 and 90 mm CO_2 Ordinates percentage saturation Abscissae oxygen pressure (After Barcroft and Poulton 1913)

With changing oxygenation it is not possible to hold both the P_{CO_2} and the pH constant (without adding acid or alkali), if P_{CO_2} remains constant the pH must shift slightly in the acid direction with increasing oxygenation since oxyhemoglobin is a stronger acid than reduced hemoglobin It is now customary to use the *standard oxygen dissociation curve* at 37° C and plasma pH of 7 40 (Fig 6 3) This has been established for human blood by careful measurements in several different laboratories, the observed values were subjected to pH and temperature corrections where necessary Severinghaus (1958) has prepared a nomogram from which the corrections can be calculated and from which dissociation curves for various values of plasma pH and temperature can be derived The reason for the sigmoid shape of the curve is discussed in section 4, chapter 4

Individual Variations and Species Differences The position of the oxygen dissociation curve is conveniently established by determining the point at which the hemoglobin is *half-saturated* On the standard curve (Fig 6 3) this occurs at an oxygen tension of 26 mm Hg The dissociation curve of the blood has been

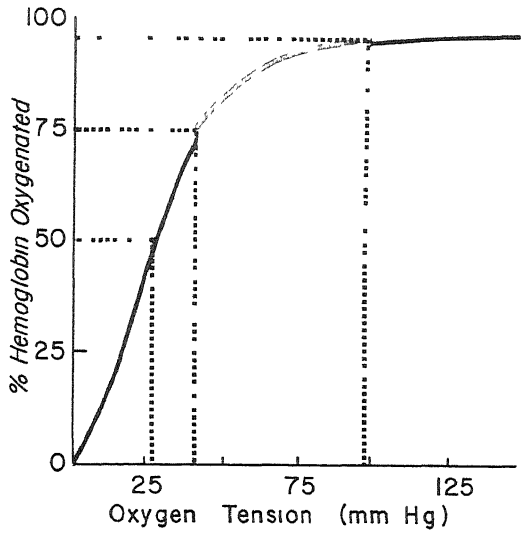


Fig 6 3 Standard oxygen dissociation curve of human blood at 37° C and plasma pH 7 40 Important points on the curve are the P_{O_2} at half-saturation (26 mm Hg) 75% saturation (40 mm Hg) and 97% saturation (99 mm Hg) The heavy line represents the normal working range between arterial and mixed venous blood It deviates slightly from the standard curve because the venous blood is 0 02 to 0 03 pH units more acid than the arterial blood

determined from time to time over a period of years in several persons and is found to be unaffected by age changes, but small variations occur in blood from different individuals In the cases of human and canine blood the tension required for half saturation at 37° C and at plasma pH 7 40 usually lies within the range of 25 to 30 mm Hg For blood from cats and sheep the value is about 1 5 times as large, indicating a lower affinity for oxygen The curve for blood of pigs, bullocks, and horses is similar to that of man Since the hemes in blood of all species are chemically identical, the variations in affinity for oxygen are attributed to differences in the globin part of the molecule The species differences in oxygen affinity are not correlated with the ability to withstand a lack of oxygen, indicating a preponderance of other factors in resistance to anoxia (Roughton, 1964)

Abnormal Hemoglobins Hemoglobin variants may modify oxygen exchange in man Hemoglobins with higher than normal affinity for oxygen occur, and are accompanied by compensatory polycythemia since they release oxygen to the tissues less readily than does normal hemoglobin There are also examples of hemoglobins with low oxygen affinity and associated anemia Alterations in the globin part of the molecule account for the variations in affinity (Lenfant et al , 1970)

Physiological Regulation of Oxygen Affinity Recent studies have revealed that the erythrocytes have an intrinsic adaptive mechanism for regulating the position of the oxygen dissociation curve (Benesch and Benesch, 1970, Lenfant et al , 1970, Rose, 1970) These cells contain a uniquely high concentration of 2,3-diphosphoglyceric acid (DPG), which tends to become reversibly bound to certain sites in the globin of reduced hemoglobin but does not bind to oxyhemoglobin The complex must dissociate for oxygen to be bound An increase in DPG therefore *shifts the oxygen dissociation curve to the right* DPG is formed by a side reaction of the glycolytic pathway, and its level is increased in response to low oxygen tension in the capillary or venous blood This adjustment requires only a day or two after ascent to a high altitude Calculations show that it would contribute to acclimatization, increasing the amount of oxygen released from the hemoglobin by possibly 10% at 15,000 ft , although at such an altitude the arterial oxygen values as well as the venous values are on a steep part of the dissociation curve In anemias and hypoxic states resulting from low cardiac output, the same adjustment occurs and, since the arterial blood is almost completely oxygenated, the shift of the curve to the right undoubtedly is an important aid in the delivery of oxygen

Alterations in the red cell DPG levels brought about by abnormal enzymatic activities are found to result in compensatory anemia or polycythemia These effects are comparable to those of hemoglobin variants previously mentioned

In blood stored for transfusions, a slow decrease in DPG brings about an increased oxygen affinity with the result that after transfusion of large quantities the release of oxygen to the tissues is subnormal This change in the blood is influenced by the preservative used, and can be retarded by the addition of inosine (Bunn et al , 1969)

Oxygen Dissociation Curve of Fetal Blood The curve for fetal blood at pH 7 40 lies to the left of the standard curve, while the curve for maternal blood is in the normal position or

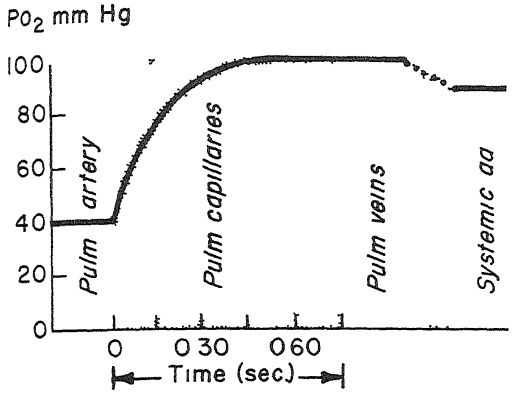


Fig 6 4 Changes in P_{O_2} of blood during its passage through the lungs (From Comroe et al 1962)

sometimes a little to the right (Darling et al , 1941) This means that the affinity for oxygen is higher in fetal than in maternal blood, a fact established for the human and several other species by various workers The difference in affinity is thought to aid in the transfer of oxygen from the maternal to the fetal circulation in the placenta Fetal hemoglobin differs in structure from adult hemoglobin, and its oxygen affinity is less influenced by DPG (Tyuma and Shimizu, 1970)

Oxygen and Carbon Dioxide Transport

OXYGENATION OF ARTERIAL BLOOD

Under normal resting conditions the blood passing through an alveolar capillary is thought to reach equilibrium with the air in the alveolus well before arriving at the end of the capillary, despite the brevity of the transit time (Fig 6 4) If this were not so, the great reductions in transit time and in venous blood oxygen content which occur during severe exercise would lead to a considerable decrease in oxygenation of the arterial blood Actually, severe exercise brings about only a very slight decrease in the oxygen tension of arterial blood while slightly increasing that of alveolar air, the alveolar-arterial P_{O_2} difference increases from 4 to 9 mm Hg under resting conditions[7] to about 13 to 18 mm Hg in severe exercise (Rossier et al , 1960)

The existence of a measurable P_{O_2} difference between alveolar air and arterial blood is attributed to two factors *1) admixture of small amounts of venous blood* with the oxygenated blood and *2) uneven effectiveness of ventilation* in relation to blood flow in different parts of the lungs Sources of venous admixture normally include the bronchial and coronary circulations, pathologically there may also be pulmonary arteriovenous shunts Blood from venous sources could decrease the arterial P_{O_2} significantly without causing a perceptible change in the P_{CO_2}, since the total difference between venous and arterial P_{CO_2} is only 6 mm Hg

Uneven ventilation relative to blood flow would not be expected to measurably influence the arterial P_{CO_2} since underventilation of some alveoli (as compared with the mean for all alveoli) would be compensated for by overventilation of others But an effect on arterial P_{O_2} occurs and is related to the shape of the oxygen dissociation curve Because of the flattening of the curve at its upper end, increasing oxygen tension above 95 mm Hg adds very little to the blood oxygen content, hence overventilated alveoli cannot supply extra oxygen sufficient to compensate for the effects of underventilated alveoli The alveolar-arterial P_{O_2} difference decreases when the alveolar oxygen tension is reduced to a level falling on a steeper part of the oxygen dissociation curve

Although the alveolar-arterial P_{O_2} difference is small and of little consequence in normal subjects, it becomes a matter of importance in certain respiratory disorders Methods have been developed to evaluate separately the effects of *1)* venous admixture, *2)* uneven ventilation-blood flow relationships, and *3)* impaired diffusion, since each of these may be a factor in respiratory dysfunction With regard to the second of these, ventilation and blood flow cannot be considered separately, the oxygen and CO_2 tensions in blood leaving the lungs are determined by the *ratio* between the ventilation of the alveoli and their capillary blood flow Alveoli that are relatively overventilated waste a fraction of their ventilation, and alveoli that are underventilated waste a fraction of their blood flow, and both types of alveoli contribute to differences in gas tensions between alveolar air and arterial blood In some disorders, ventilation blood flow ratios are uneven enough to cause anoxemia or to exacerbate anoxemia resulting from decreased alveolar

[7] Slightly higher values are found by many investigators, the method used for sampling the alveolar air is of critical importance

ventilation (Comroe et al , 1962, Rahn and Fahri, 1964, West, 1965)

The Working Range of Blood Oxygenation The normal range between arterial and mixed venous blood under resting conditions is shown in Fig 6 3 In arterial blood the hemoglobin is 96 to 97% saturated with oxygen In the tissue capillaries, outward diffusion of dissolved oxygen reduces the blood oxygen tension Accordingly, oxyhemoglobin dissociates, and the released oxygen diffuses away in the dissolved form When the tension has been reduced to 40 mm Hg, the level of saturation is about 75% A slight decrease in the blood pH occurs (0 02 to 0 03 pH units) due to the uptake of carbon dioxide, hence the physiological curve deviates slightly from the standard dissociation curve at pH 7 40

Depending upon the particular tissue and its activity at the moment, the blood in its passage through the capillaries loses from one-tenth to nearly all of its oxygen load The figure used to express the proportion of the total oxygen content given up to the tissue is called the *coefficient of utilization* Thus if the oxygen content of arterial blood is 19 5 volumes per cent and that of venous blood is 15 volumes per cent (i e , an arteriovenous difference of 4 5 volumes per cent) the coefficient is 4 5/19 5 or 0 23 As indicated above, the coefficient varies considerably for different tissues and for the same tissue in accordance with blood flow and metabolic activity During heavy exercise the overall coefficient of utilization increases to about 0 6, and that of active muscle to as high as 0 8

Certain features associated with the shape and position of the oxygen dissociation curve are of the utmost physiological importance At the normal oxygen tension of alveolar air at sea level, the hemoglobin in the blood passing through the lungs becomes almost completely saturated The flattening of the upper part of the curve means that relatively little reduction in the per cent saturation occurs until the alveolar Po_2 falls to about 60% of its value at sea level Such a decrease would occur on ascending to an altitude of about 10,000 ft , and at sea level is not uncommon in cardiopulmonary disease Within this range, wide variations in alveolar Po_2 can occur with minimal changes in the total oxygen load of arterial blood. This circumstance constitutes a considerable safety factor On the lower part of the curve, the steep slope means that in this region a given fall in oxygen tension causes the release of a relatively large volume of oxygen The relation between oxygen content and oxygen tension therefore is such as to favor a nearly maximum *uptake* of oxygen in the lungs as long as the tension is above 60 mm Hg and an adequate *release* of the gas at the lower tensions which prevail in the tissues The slope of the curve would favor release at still lower tensions, but at some point the diffusion gradient would become insufficient to support the rapid transfer of oxygen into the cells

The effects of increasing hydrogen ion concentration and Pco_2 on oxygen dissociation (shifting the curve to the right) tend to favor the release of oxygen to the tissues since carbon dioxide and sometimes lactic acid are produced by tissue activity A slight increase in temperature, such as may occur in working muscle, has a similar effect However, these shifts in the position of the curve are small and in working muscle are much less important than the great increase that occurs in the blood flow The influence of temperature is of some importance in hyperthermia, and requires particular consideration when patients are subjected to hypothermia for surgical procedures Estimates of oxygen tension from oximeter readings taken during hypothermia would be higher than the true values if the effect of temperature were disregarded

OXYGEN CONTENT–Po_2 CURVES

When *per cent saturation* is plotted against oxygen tension, as in Fig 6 3, the importance of the hemoglobin concentration is not evident, the curve for blood from an anemic patient appears the same as that for normal blood The difference becomes apparent if oxygen *content* is plotted against oxygen tension, as in Figure 6 5 Such a curve for blood of any hemoglobin concentration can be obtained from the standard curve by calculating the volumes per cent of oxygen combined with hemoglobin (per cent saturation × oxygen capacity) at each of a series of Po_2 values and adding the small concentration of dissolved oxygen present at that level It is evident from Figure 6.5 that a 50% decrease in hemoglobin, such as might occur in anemia, would halve the amount of oxygen delivered by 100 ml of blood if the working range were unaltered Partial compensation occurs by an increase in the coefficient of utilization, which in the diagram extends the working

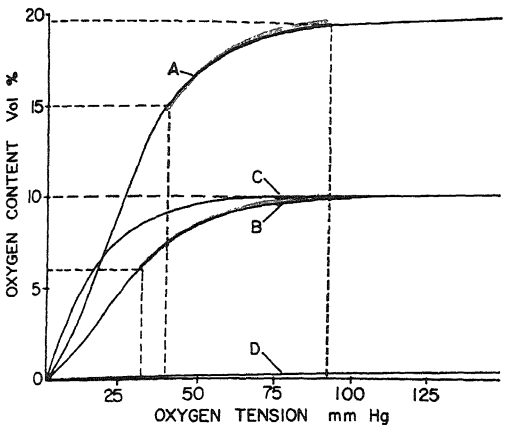


Fig 6 5 Effects of anemia and of carbon monoxide on the oxygen dissociation curve of blood at pH 7 40 and 37°C *A*) Normal blood hemoglobin 15 gm/100 ml *B*) Anemia hemoglobin 7 5 gm/100 ml *C*) Carbon monoxide poisoning hemoglobin 15 gm/100 ml half-saturated with CO *D*) Curve representing oxygen in physical solution Heavy lines indicate the working range in normal and anemic subjects

range downward to a tension of 32 mm Hg, but the oxygen delivered by 100 ml of blood is still not as much as in the normal subject Not shown in Figure 6 5 is a small shift of the anemic curve to the right induced by an increase in 2,3-diphosphoglyceric acid in the red cells Additional compensation is achieved by a rise in the cardiac output, which increases the oxygen delivery by increasing the volume flow of blood through the tissues The circulatory adjustments are aided by a decrease in blood viscosity (a consequence of erythrocyte depletion), which reduces the peripheral resistance

Effects of Carbon Monoxide Hemoglobin and carbon monoxide react reversibly to form *carboxyhemoglobin* As with oxygen, the hemoglobin becomes saturated in the proportion of one molecule of gas to one atom of iron, and the physicochemical aspects of the two reactions are very similar except that the affinity of hemoglobin for carbon monoxide is about 200 times its affinity for oxygen When both gases are present they compete for the hemoglobin

The presence of 0 1% of carbon monoxide in the inspired air over a period of a few hours (or a higher concentration for a shorter time) is sufficient to convert 50% of the hemoglobin in the blood to carboxyhemoglobin, and at about this level the effect becomes lethal By contrast, an anemic patient whose blood hemoglobin is reduced to 50% of normal shows little evidence of anoxia under resting conditions The major reason for this difference is shown in Figure 6 5 It will be seen that the position of the oxygen dissociation curve is shifted to the left by carbon monoxide, with the result that the ability to delivery oxygen to the tissues is greatly reduced In other words, as the per cent of carboxyhemoglobin increases, the remaining active hemoglobin holds more tightly to the oxygen with which it combines This is explained as follows in the hemoglobin molecule there are four hemes, each containing one atom of iron, and when one or more of the hemes is combined with carbon monoxide an interaction takes place which increases the affinity of the remaining hemes for oxygen (Roughton, 1964)

VARIATION OF BLOOD CO_2 CONTENT WITH P_{CO_2}

Figure 6 6 shows the manner in which the CO_2 *content* of the blood varies when the blood is exposed to gas mixtures of varying CO_2 *tension* In the physiological range, an increase in P_{CO_2} induces an almost proportional increase in CO_2 content, i e , the curve is nearly a straight line in this region A pH decrease, calculable from the Henderson-Hasselbalch equation, accompanies a P_{CO_2} increase but is very small in this range The exact position of the CO_2 content-P_{CO_2} curve depends upon the total buffering power of the blood Christiansen, Douglas, and Haldane (1914) first noted that the curve is higher for reduced blood than for oxygenated blood In other words, reduced blood has more buffering power than oxygenated blood and can carry more carbon dioxide at a given P_{CO_2} This suggested that oxyhemoglobin may behave as a stronger acid than reduced hemoglobin Thus oxygenation would cause the hemoglobin to release hydrogen ions, which would then be neutralized and would use up some of the buffer base (including some bicarbonate), and accordingly the position of the CO_2 content-P_{CO_2} curve would be shifted downward Titration curves for oxygenated and reduced hemoglobin supported this interpretation However, this mechanism is now known to be only a part of the explanation for the shift of the curve The other factor is the effect of oxygenation in diminishing the power of hemoglobin to combine directly with carbon dioxide

Transport of CO_2 as Carbaminohemoglobin Carbon dioxide combines with hemoglobin, not

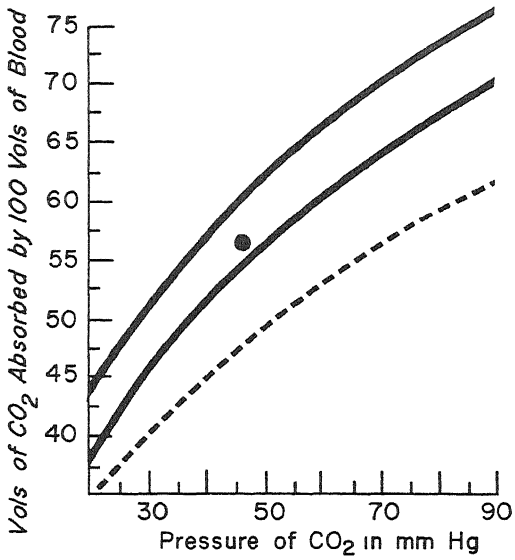


Fig 6 6 Carbon dioxide dissociation curves of fully reduced human blood (upper solid line) in presence of hydrogen and fully oxygenated human blood (lower solid line) in presence of air Volumes of CO_2 along the ordinate CO_2 tension along the abscissa Line AVB is the so-called physiological dissociation curve of CO_2 i e as a result of the reduction of hemoglobin the relation of volumes of CO_2 to CO_2 tension is indicated at points along this line and not along the lower curve for oxygenated blood At *A* (arterial point) are indicated the volume and tension of CO_2 in arterial blood Point *B* indicates the conditions in fully reduced blood Point *V* (venous point) represents the degree of reduction of hemoglobin which normally occurs in the body during the passage of the blood through the systemic capillaries The interrupted line below is the CO_2 dissociation curve for oxygenated dog s blood (Modified from Christiansen Douglas and Haldane 1914)

with the hemes as do O_2 and CO, but rather with free amino groups in the globin part of the molecule The reaction can be represented as follows

$$Hb—NH_2 + CO_2 \rightleftharpoons Hb—NHCOO^- + H^+$$

This occurs at a very rapid rate, requiring no enzyme, and the CO_2 does not go through the stage of H_2CO_3 The carbamic acid is almost completely ionized, but the pH change is minimized by the buffering power of the hemoglobin (which increases as the hemoglobin is desaturated, and vice versa)

Carbamino compounds of CO_2 with amino acids are well known, and the technique of estimating them was adapted to the measurement of carbaminohemoglobin by Ferguson and Roughton (1934) Hemoglobin solutions containing bicarbonate were equilibrated with suitable tensions of CO_2 and then were mixed very rapidly with alkaline barium chloride The alkali stabilized the carbamino compound, probably by inducing its complete ionization, and also converted the dissolved CO_2 and bicarbonate to carbonate which was then precipitated by the barium chloride The $BaCO_3$ precipitate was removed by centrifuging The carbaminohemoglobin, which remained in solution, could then be decomposed by acidification and the released CO_2 measured by the Van Slyke technique It was found that *oxygenation of the hemoglobin* greatly reduced the amount of carbamino-bound CO_2 Although an increase in P_{CO_2} would be expected to favor the formation of carbaminohemoglobin, a decrease in pH has an opposite effect (by converting $Hb—NH_2$ to $Hb—NH_3^+$, which does not react with CO_2), hence the dissociation curve of carbaminohemoglobin is practically flat over the physiological range, and the variation in P_{CO_2} between arterial and venous levels has little effect on the carbamino reaction The chief factor influencing the formation and breakdown of carbaminohemoglobin in the body is the variation in oxygenation It has been estimated that 20 to 30% of the carbon dioxide delivered to the alveolar air under resting conditions is transported as carbaminohemoglobin, although only 6 to 7% of the CO_2 content of the venous blood is in this form (Davenport, 1958, Roughton, 1964, Rossi-Bernardi and Roughton, 1967)

Only traces of carbamino compounds are formed by the plasma proteins, and the concentrations do not vary significantly between arterial and venous blood

Transport of CO_2 as Bicarbonate Carbon dioxide produced in the tissues diffuses into the blood in the capillaries as free CO_2 in solution In the blood, a large part of it is converted to carbonic acid·

$$CO_2 + H_2O \rightleftharpoons H_2CO_3 \qquad (1)$$

This reaction will take place in the absence of a catalyst, but the uncatalyzed reaction is much too slow to account for the rapid exchanges which take place in the body The reaction rate is accelerated enormously (in both directions) by an enzyme called *carbonic anhydrase*, which is found inside the erythrocytes but not in the

blood plasma [8] The formation and decomposition of H_2CO_3 therefore occur primarily in the red cells

The equilibrium point of reaction (1) lies far to the left However, the hydration of CO_2 is followed by the rapid dissociation of carbonic acid

$$CO_2 + H_2O \rightleftharpoons H_2CO_3 \rightleftharpoons H^+ HCO_3^-$$

It is obvious that, by the law of mass action, an increase in P_{CO_2} and the removal of hydrogen ions would tend to drive the two reactions to the right This is what happens as the blood is passing through the *tissue capillaries* CO_2 diffusing into the blood increases the P_{CO_2}, and hydrogen ions are removed by the buffers, especially by hemoglobin since oxyhemoglobin is being converted to the reduced form, a weaker (less dissociated) acid The opposite changes occur as the blood passes through the *alveolar capillaries* the diffusion of CO_2 out of the blood reduces the P_{CO_2}, and the hemoglobin becomes oxygenated and releases hydrogen ions, with the result that the equilibrium is shifted to the left

These interlocking reactions can be formulated as follows

$$CO_2 + H_2O \rightleftharpoons H^+ + HCO_3^-$$

$$H^+ + HbO_2^- \rightleftharpoons HHb + O_2$$

[8] Carbonic anhydrase is a protein of comparatively low molecular weight (about 30,000) containing one atom of zinc per molecule It is present also in tissues such as gastric and intestinal mucosa, renal cortex, pancreas, brain, spleen, red muscle, liver, and salivary glands, and in the oviducts of hens where its function appears to be concerned in the deposition of calcium in the egg shell Its physiological significance in most of these situations is unknown The enzyme is inhibited by sulfanilamide, cyanide, heavy metals, sulfocyanate, and certain other ions Different sulfonamide drugs differ in their inhibitory potency The ones most commonly used therapeutically are weak inhibitors, but some have been developed for this special purpose Among these is acetazolamide (Diamox), which is used as a diuretic in cardiac insufficiency It acts on the kidney, decreasing the tubular reabsorption of bicarbonate and thus increasing the excretion of sodium and potassium as bicarbonates It also affects the brain (perhaps indirectly), and is used as an anticonvulsant In experimental animals, large doses of Diamox initially induce a fall in CO_2 output, with recovery to the normal rate in about an hour The compensation is brought about by a 40% increase in the P_{CO_2} of the venous blood and a doubling of the ventilation rate (Roughton, 1964)

Here oxyhemoglobin is represented as an ionized acid and reduced hemoglobin as unionized It must be recognized that neither is *completely* ionized or unionized, this formulation merely symbolizes the fact that oxyhemoglobin is the stronger acid of the two When the equilibrium shifts, hydrogen ions released by one reaction can be taken up by the other The overall equilibrium (the sum of the two reactions above) therefore is

$$CO_2 + H_2O + HbO_2^- \underset{\text{lungs}}{\overset{\text{tissues}}{\rightleftharpoons}} HCO_3^- + HHb + O_2$$

This is called the isohydric shift The reaction does not go to completion in either direction, but the equilibrium is shifted to the right by conditions in the tissues and to the left by conditions in the lungs Thus it can be said that CO_2 tends to drive out oxygen, and oxygen tends to drive out CO_2 The shift occurs with very little change in the plasma pH

The negative ions, HbO_2^- on the left and HCO_3^- on the right, are balanced by the presence of positive ions (mostly K^+) inside the red cells [9] However, the cations take no active part in the reaction, and the common statement that they have been "captured" by oxyhemoglobin on the one hand or by bicarbonate on the other is likely to introduce confusion It would be more accurate to consider that *hydrogen ions* are captured either by reduced hemoglobin or by the bicarbonate system

The "heme-linked" or "oxylabile" ionizing groups, i e , those that dissociate to release hydrogen ions more readily in oxygenated than in reduced hemoglobin, are believed to be imidazole groups of the amino acid histidine, a constituent of the globin part of the molecule

An over-simplification must be recognized in the above formulation The hemoglobin molecule contains 4 hemes (sect 4, chap 4) and each heme appears to be associated with one ionizable hydrogen Therefore, fully oxygenated hemoglobin should be represented as $H_4Hb_4(O_2)_4$, and there are intermediate forms such as $H_4Hb_4(O_2)_3$, etc However, the simpler formulation is qualitatively satisfactory if the

[9] The cation content frequently is referred to as "base," but in modern acid-base theory the term "base" is reserved for substances able to accept protons (sect 4, chap 11)

symbol Hb is understood to represent one heme and not a hemoglobin molecule

The Chloride-Bicarbonate Shift Chloride and bicarbonate ions pass freely through the red cell membrane, while cations traverse it only very slowly When the blood is in the tissue capillaries bicarbonate ions formed inside the red cells tend to diffuse out to the plasma Since they cannot be accompanied by cations the red cells are left with a positive electric charge which tends to draw anions into the cells Hence chloride ions, being the most abundant anions, migrate inward in exchange for bicarbonate The reverse movements occur when the blood is passing through the lungs and intracellular bicarbonate is being converted to CO_2 The intracellular-extracellular distribution of ions is described by Donnan's theory, which requires the following relations

$$\frac{[HCO_3^-]_{cells}}{[HCO_3^-]_{plasma}} = \frac{[Cl^-]_{cells}}{[Cl^-]_{plasma}} = \frac{[H^+]_{plasma}}{[H^+]_{cells}}$$

This "Donnan ratio" varies from about 0 72 in arterial blood to 0 75 in venous blood (Davenport, 1958) It follows that *1)* the greater portion of the bicarbonate is carried in the plasma, and *2)* the pH inside the cells is lower than that of the plasma by about 0 12 to 0 14 pH units With the redistribution of ions occurring between the arterial side and the venous side of the circulatory system there is an alteration of osmotic relationships such that some water enters the cells along with the chloride, and a slight swelling of the cells occurs

The exchanges occurring as the blood passes through the lungs are summarized in Roughton's diagram (Fig 6 7)

Gas Tensions in the Tissues

Oxygen Tension In the tissues, oxygen is being supplied by diffusion and consumed within the cells, hence diffusion gradients are present and a uniform Po_2 cannot exist At any given point within a tissue the Po_2 level represents a dynamic balance between oxygen supply and utilization The rate of *supply* is influenced by the following factors arterial Po_2 and oxygen capacity, blood flow, distance from a capillary, distance between capillaries, distance along the capillary, radius of the capillary, and the rate of diffusion of oxygen in the tissue The rate of oxygen *utilization* normally depends upon the energy requirement of the tissue, it varies with activity but is not limited by the availability of oxygen except when Po_2 is reduced to a very low level

Davies and Bronk (1957) used a microelectrode for the polarographic measurement of Po_2 at the surface of the cerebral cortex in cats, and were able to demonstrate the existence of steep downward diffusion gradients as the electrode was moved away from an arteriole or a venule Others have used modifications of this technique to record Po_2 in various parts of the brain and in other tissues (Payne and Hill, 1966)

The distribution of a gas in a tissue can be analyzed theoretically and its *average* tension estimated by calculation with the aid of a diffusion model originally suggested by Krogh It is assumed that the tissue is functionally divided into cylindrical segments, each supplied with a centrally situated capillary (Kety, 1957), Gleichmann et al (1962) applied this reasoning to estimate the mean oxygen tension in the cerebral cortex of the dog, and compared the results with polarographic measurements of Po_2 at the surface of the cortex The active surface of each electrode covered an area 2 to 4 mm in diameter, and large vessels were absent from the areas chosen The surface values agreed fairly well with those calculated for mean Po_2 in the tissue, and usually came close to the venous blood Po_2 The observed values were 25 to 42 mm Hg in animals with ventilation rates approximately normal

Other workers have reported lower Po_2 values from the surface of the cortex and from electrodes thrust into various parts of the brain For example, Meyer and Gotoh (1961) reported

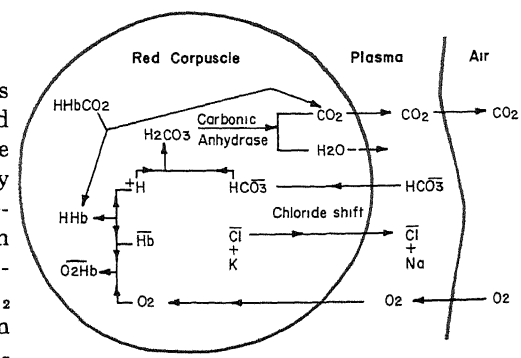


Fig 6 7 Showing main processes occurring in the blood during the output of carbon dioxide and uptake of oxygen in the lung (After Roughton 1961)

surface values of 10 to 20 mm Hg in cats and monkeys Cater et al (1961) made an extensive exploration of the rat brain, finding values of 33 to 60 mm Hg at the cortical surface, 20 to 38 in gray matter of the cortex, 6 to 16 in white matter of the cortex, and 10 to 15 in hypothalamic nuclei In rat muscle the values were 14 to 26, and in subcutaneous tissue there were wide variations from 8 to 62 Jamieson and Van den Brenk (1964) reported the following mean values for rat tissues cerebral cortex 34, liver 13, kidney 17, and spleen 17 mm Hg

In the various tissues studied, rapid variations in Po_2 occur in relation to metabolic, circulatory and respiratory changes induced by various means

CO_2 Tension Diffusion gradients of Pco_2 must be assumed to exist between the tissue cells and the blood in the capillaries With the aid of Krogh's diffusion model, Gleichmann et al (1962) calculated the *average* Pco_2 in the cerebral cortex of the dog in each of a series of experiments A statistical analysis of the data indicated that the average tissue Pco_2 is only about 0 7 mm Hg higher than the mid-capillary blood Pco_2 (i e , the average of arterial and cerebral venous values) CO_2 electrodes placed on the surface of the cortex gave values which in each experiment agreed closely with the estimated average tissue Pco_2 The values ranged from 31 to 58 mm Hg Meyer and Gotoh (1961) found that the Pco_2 of the cerebral cortex fluctuates rapidly in relation to metabolic, circulatory, and respiratory alterations

Adjustments in Metabolic Requirements

Increased activity of any tissue entails an increase in oxygen consumption and CO_2 production On the other hand, an increase in the oxygen supply above that which is adequate does not increase the oxygen usage There are two ways in which an increased demand for oxygen may be met *1*) by increasing the blood flow through the tissue, and *2*) by raising the coefficient of oxygen utilization Both of these factors come into play, but the extent to which each operates is different for different tissues The coefficient of oxygen utilization is increased by establishing a steeper Po_2 gradient between the capillary blood and the tissue cells, this is brought about by *1*) a decrease in the intracellular Po_2 as utilization increases, *2*) a shortening of the distance through which oxygen must diffuse, i e , the opening up of more capillaries so that the radius of the cylinder of tissue supplied by each capillary is reduced, and *3*) increased dissociation of oxyhemoglobin due to increased production of acids (carbonic and lactic) and a rise in temperature (Otis, 1963)

Total, Dead Space and Alveolar Ventilations

The lung is made up of tubes for airflow and alveoli for exchange of gases between air and blood At the end of a normal expiration, the airways and alveoli are filled with alveolar gas At the end of a normal inspiration, however, the airways are filled with moist room air and the balance of the inspired volume has been added to the volume of the alveolar spaces During the subsequent expiration, room air is washed out of the airways by alveolar air and the cycle begins again Therefore one expired breath (tidal volume) contains gas from two compartments, one filled with room air containing a negligible concentration of carbon dioxide, and the other filled with alveolar air with a partial pressure of carbon dioxide (P_{CO_2}) in equilibrium with arterial blood

The volume of the airways of an adult man is about 150 ml At a tidal volume of 500 ml and a respiratory frequency of 12/min , the total ventilation is 6 0 liters/min of which the airway ventilation is $150 \times 12 = 1\,8$ liters/min and the alveolar ventilation is $350 \times 12 = 4\,2$ liters/min It is customary to refer to airway ventilation as "dead space" ventilation because it is not associated with exchange of gases with capillary blood The terms "anatomical" and "physiological" dead space ventilation are used to distinguish simple airway ventilation from ventilation of airways together with alveolar spaces which do not participate in gas exchange A lung lacking its pulmonary artery, for example, would have only dead space ventilation despite its normal anatomical subdivision into airways and alveoli

Anatomical dead space can be measured postmortem by making casts of the airways from mouth to terminal bronchioles and measuring their volume It can also be measured approximately by the technique described in Figure 6.8 (Fowler, 1952) This requires measuring volume of expired air and nitrogen concentration in expired air simultaneously A subject breathing room air (79% nitrogen) takes in a breath of 100% oxygen and then exhales slowly The nitrogen meter records 0% nitrogen during the period of inspiration and during the first part of expiration as the dead space air is washed out The meter then samples a mixture of rising nitrogen concentration as dead space gas is mixed with alveolar gas Ideally, if the inspired oxygen is evenly distributed throughout the alveoli, the meter then records an unvarying nitrogen concentration The measurement of dead space is calculated to be the volume expired up to the time at which one-half the volume of the rising phase has been delivered Measurements so made in normal subjects indicate that anatomic dead space at the end of a maximal expiration is about one-half its value at the end of a complete inspiration This indicates the extent to which the airway volume is increased by elongation and widening as traction of airways by lung parenchyma increases with lung volume

Total ("anatomical" plus "physiological") dead space ventilation is readily measured provided the carbon dioxide content of alveolar air is known This can be measured by sampling alveolar gas as described in a previous chapter Since alveolar air is normally in equilibrium with arterial blood, the alveolar partial pressure can also be obtained from the arterial partial pressure Mixed expired air is a mixture of room air from the airways and nonperfused alveoli together with air from perfused alveoli Its carbon dioxide partial pressure is lower in proportion to the dilution of the alveolar fraction, in equilibrium with capillary blood, by the dead space fraction, containing virtually no carbon dioxide If, for example, the arterial P_{CO_2} is 40 mm Hg and the partial pressure of carbon dioxide in mixed expired air ($P_{E_{CO_2}}$) is 30 mm Hg, it must be concluded that alveolar gas was diluted one part room air (dead space) in three parts alveolar air

The equation for calculating the ratio dead space volume/tidal volume is

$$V_D/V_T = (P_{A_{CO_2}} - P_{E_{CO_2}})/P_{A_{CO_2}}$$

where V_D—dead space volume, V_T—tidal vol-

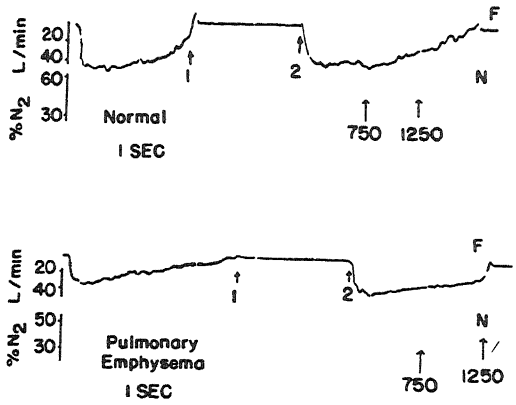


Fig 6 8 Single breath nitrogen dilution method for 1) measuring anatomic dead space and 2) detecting nonuniform distribution of inspired air F = respiratory flow N = nitrogen concentration Horizontal axis is time Integration of F would give volume Expiration after breathing room air is indicated to left of 1 Between 1 and 2 patient inspired one breath of oxygen and at 2 expiration began See text for description of nitrogen curve and its use for measuring anatomical dead space Nonuniform distribution of inspired oxygen is shown by continued increase of nitrogen concentration after rapidly rising phase The increase in concentration between expired volume of 750 and 1250 is 1% in the normal subject and 9% in the patient with emphysema indicating that mixing within the lung is less complete in the latter (From Comroe Jr and Fowler 1951)

ume, $P_{A_{CO}}$—the partial pressure of carbon dioxide in alveolar gas or arterial blood and $P_{E_{CO}}$ —the partial pressure of carbon dioxide in mixed expired air In the instance cited,

$$V_D/V_T = (40 - 30)/40 = 0\ 25$$

If the minute ventilation were 6 liters/min and the respiratory frequency 12/min, the tidal volume would be 500 ml and the dead space volume 500 × 0 25 = 125 ml. Dead space ventilation would be 125 × 12 = 1500 ml /min and the balance of 4½ liters/min would be alveolar ventilation V_D/V_T does not exceed 0 30 in normal subjects

The design of the lung is such that dead space ventilation is obligatory, i e, the proportion of a tidal volume delivered to the alveoli is whatever exceeds the volume of the dead space Table 6 5 shows how alveolar ventilation is affected by tidal volume when total ventilation and dead space volume are kept constant It shows that dead space ventilation increases at the expense of alveolar ventilation as tidal volume approaches dead space volume It is clear from this table that any increase in dead space, such as physiological dead space or anatomical dead space in the form of breathing equipment, decreases alveolar ventilation unless total ventilation is increased The patient with marginal ventilatory ability is therefore disadvantaged by any increase in dead space and may be assisted by decreasing his anatomical dead space Tracheostomy reduces the anatomical dead space about 50% by eliminating ventilation of the mouth, larynx, and pharynx One of its favorable aspects is the increase in alveolar ventilation which results from reducing dead space

Dead space ventilation is not altogether useless Inspired air is warmed, saturated with water vapor at 37° C and filtered Panting, which is primarily dead space ventilation, is an important temperature regulating mechanism in some animals The vital functions of ventilation, however, concern maintaining a sufficiently high partial pressure of oxygen in the alveoli and excreting carbon dioxide These are accomplished exclusively by alveolar ventilation

The Alveolar-Arterial Partial Pressure Difference for Oxygen

Alveolar air contains four gases (O_2, CO_2, N_2, H_2O) the sum of whose partial pressures equals the barometric pressure (see previous chapters) The partial pressures of two of these, N_2 and H_2O, is fixed, for the pressure of water vapor in saturated air at body temperature (37° C) is a constant (47 mm Hg) and the nitrogen inspired into the lung is neither taken into the blood stream nor is it added to from the blood stream The sum of the partial pressures of oxygen and carbon dioxide is therefore also a constant, approximately 150 mm Hg at sea level This means that the subject breathing room air can vary his alveolar P_{O_2} only by varying his alveolar P_{CO_2}

Alveolar P_{CO_2} is directly proportional to the

Table 6 5

Frequency per min	V_T *ml*	V_D *ml*	Ventilation (l /min)		
			Total	Dead Space	Alveolar
10	600	150	6 00	1 50	4 50
15	400	150	6 00	2 25	3 75
20	400	150	6 00	3 00	3 00
30	200	150	6 00	4 50	1 50
40	150	150	6 00	6 00	0

rate at which carbon dioxide is produced by the tissues and inversely proportional to alveolar ventilation

$$P_{A_{CO}} \cong \frac{\text{Production}}{\text{Excretion}} \cong \frac{\text{Metabolic rate}}{\text{Alveolar ventilation}}$$

Clearly, therefore, at any metabolic rate, alveolar ventilation is the key to alveolar P_{O_2} Insufficient alveolar ventilation causes hypoxia because carbon dioxide is produced more rapidly than it is excreted, alveolar P_{CO_2} rises and alveolar P_{O_2} must consequently fall Conversely, increasing alveolar ventilation at a time when the rate of CO_2 production is constant lowers alveolar P_{CO_2} resulting in an increase in alveolar P_{O_2}

Alveolar P_{O_2} ($P_{A_{O_2}}$) can be varied also by changing the total barometric pressure when the oxygen concentration (F) is kept constant (normally 0 2093), by changing the concentration in inspired air ($F_{I_{O_2}}$) or by changes in both pressure and concentration For example, the Himalayan mountain climber can compensate for the fall in $P_{A_{O_2}}$ resulting from low barometric pressure by breathing mixtures in which oxygen concentration is increased at the expense of nitrogen concentration

These factors upon which $P_{A_{O_2}}$ depend are included in the alveolar air equation, one form of which is as follows

$$P_{A_{O_2}} = (BP - 47)F_{I_{O_2}} - (P_{A_{CO}}/R) + [F_{I_{O_2}} P_{A_{CO}} (1 - R)/R]$$

For the derivation of this and other equations used in respiratory physiology, the interested student is referred to Rahn and Fenn (1955) or Comroe (1965) The critical terms in the equation are the barometric pressure (BP), the concentration of oxygen in inspired air (F_{I_O}) and the partial pressure of carbon dioxide in alveolar gas ($P_{A_{CO_2}}$) or arterial blood The final term

$$[F_{I_O}, P_{A_{CO}} (1 - R)/R]$$

is numerically small and is 0 when the gas exchange ratio or respiratory quotient (R) is 1 It should be clear from substituting normal values for $P_{A_{CO}}$ (40 mm Hg) and R (0 8) that $P_{A_{O_2}}$ can be increased to about 660 mm Hg at sea level (760 mm Hg) by breathing 100% oxygen and to more than one atmosphere by placing the subject in a pressure chamber where the barometric pressure can be raised to two or three atmospheres (hyperbaric oxygen chamber). $P_{A_{O_2}}$ can be raised relatively little by increasing alveolar ventilation, for normal $P_{A_{CO_2}}$ is only 40 mm Hg and $P_{A_{CO_2}}$ is rarely lowered to less than 20 mm Hg during hyperventilation On the other hand, CO_2 retention from alveolar hypoventilation can be sufficiently severe to double normal $P_{A_{CO}}$ thereby reducing P_{A_O} substantially Under normal circumstances, however, the regulation of respiration is such that increases in CO_2 production are attended by commensurate increases in alveolar ventilation, and a normal $P_{A_{CO}}$ and normal P_{A_O} are maintained

Alveolar P_{O_2}

The partial pressure of oxygen in alveolar gas ($P_{A_{O_2}}$) calculated from the alveolar equation is a mean value representative of all the alveoli This value is invariably higher than the partial pressure of oxygen in arterial blood (P_{A_O}) by about 13 mm Hg in normal individuals and by considerably greater amounts in certain abnormalities The reasons for this difference in oxygen tensions are as follows (Fig 6 9)

ANATOMICAL SHUNTS

Some blood passes from the venous to the arterial circulation without being exposed to the P_{A_O} calculated by the alveolar equation (Fig 6 9) Some of this blood bypasses the lung entirely, the best known examples being blood from the coronary circulation draining directly into the left heart through the Thebesian veins,

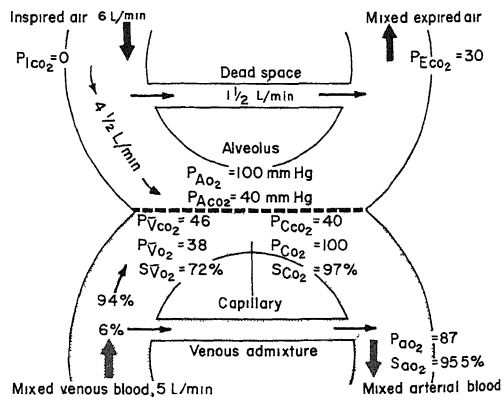


Fig 6 9 Idealized schema of lung Upper half represents ventilation lower half perfusion See text for meaning of symbols and concepts $P_{C_{O_2}}$, $P_{C_{O_2}}$, and $S_{C_{O_2}}$ are partial pressures of carbon dioxide and oxygen at the end of the pulmonary capillary and saturation of blood at the end of the capillary respectively

and venous blood in the bronchial veins draining into the pulmonary circulation distal to the capillary bed It is estimated that less than 6% of the cardiac output follows these anatomic shunts Clearly, they result in lowering the P_{O_2} of arterial blood below the P_{O_2} of pulmonary capillary blood in equilibrium with alveolar gas

It is instructive to calculate the magnitude of this effect under specified conditions Assuming, for example, that P_{A_O} equals 100 mm Hg, the oxyhemoglobin dissociation curve indicates that pulmonary capillary blood in equilibrium with this gas has an oxyhemoglobin saturation of 97% If the capillary-venous difference in saturation is 25%, a common value, the venous saturation will be 72% If venous shunts account for 6% of the cardiac output, leaving 94% to pass through the pulmonary capillaries, the resultant mixture will have a saturation of (0 06 × 72) + (0 94 × 97) = 4 32 + 91 18 = 95 50% The partial pressure of oxygen at this saturation is 87 mm Hg, 13 mm Hg lower than in pulmonary capillary blood There is normally, therefore, a difference between pulmonary capillary P_{O_2} (P_{c_O}) and arterial P_{O_2} (P_{a_O}) related to the magnitude of the venous admixture from anatomic shunts

PHYSIOLOGICAL SHUNTS

The lung has several million alveoli Because of normal regional differences in compliances and airway resistances, variations in ventilation inevitably result, producing localized differences in $P_{A_{CO}}$ and consequently in $P_{A_{O_2}}$ The ultimate example of physiological shunting is circulation through the capillaries of nonventilated alveoli, the venous admixture effect of such blood comparing to that of blood coursing through anatomical shunts

One of the interesting effects of regional differences in ventilation and $P_{A_{O_2}}$ is that $P_{a_{O_2}}$ is lowered even if hypoventilation of some alveoli is matched by hyperventilation of others This effect results from the nonlinearity of oxyhemoglobin dissociation Consider, for example, arbitrarily dividing the pulmonary blood flow into three equal streams coursing through capillaries in equilibrium with alveolar oxygen partial pressures of 60, 80, and 100 mm Hg One might assume that the mixed blood collected from these would have a P_{O_2} of 80 mm Hg, i e., (0 333 × 60) + (0 333 × 80) + (0 333 × 100) = 80 The correct P_{O_2}, however, would be found by calculating the P_{O_2} at the saturation of the mixed blood This calculation is performed as follows The dissociation curve indicates that saturations at P_{O_2} of 60, 80, and 100 mm Hg are 89%, 94%, and 97%, respectively The saturation of the mixture, therefore, would be (0 333 × 89) + (0 333 × 94) + (0 333 × 97) = 29 6 + 31 3 + 32 3 = 93 2% The P_{O_2} at a saturation of 93 2% is 75 mm Hg, rather than 80 mm Hg It is clear, therefore, that a lung characterized by large variations in P_{A_O} inevitably delivers a blood mixture which is hypoxic relative to mean P_{A_O}

The concept of anatomical and physiological venous admixture is analogous to the concept of anatomical and physiological dead space (Fig 6 9) The difference $P_{c_O} - P_{a_O}$ is a measure of venous admixture just as the difference $P_{A_{CO}} - P_{E_{CO}}$ is a measure of the admixture of dead space ventilation to alveolar ventilation (Riley and Cournand, 1949 and 1951, Riley et al , 1951) The fraction of the cardiac output which is venous admixture (Q_{VA}/Q_T) is proportional to the venous admixture effect on the saturation of capillary blood ($Sc_O - Sa_O$) divided by the total saturation difference between venous and capillary blood ($Sc_{O_2} - S\bar{v}_{O_2}$)

$$Q_{VA}/Q_T = (Sc_O - Sa_O)/(Sc_O - S\bar{v}_O)$$

In Figure 6 9, for example, Q_{VA}/Q_T = 97 − 95 5/97 − 72 = 1 5/25 = 0 06 This equation can be applied to any situation in which one wishes to measure relative flows of blood through two branches of a confluent system It finds application clinically in calculating flows through the anatomical defects of children with congenital heart disease

Alveoli can be classified by their ventilation-perfusion ratios, those with a high value contributing to the physiologic dead space and those with a low value contributing to the venous admixture-like effect of physiological shunts Inasmuch as normal alveolar ventilation is about 4 liters/min and cardiac output about 5 liters/min , a normal ventilation-perfusion ratio for the entire lung is approximately 0 8

Measuring dead space ventilation and measuring venous admixture is one means of detecting ventilation-perfusion abnormalities There are other means Nonuniform distribution of inspired air, for example, shows up on the nitrogen meter record of Figure 6 8 as a progressive increase in nitrogen concentration with time as the parts of the lung which are

most poorly ventilated empty Another popular method is to measure the concentration of nitrogen in alveolar air after the patient has breathed nitrogen-free oxygen for a specified period Normal individuals have less than 2 5% nitrogen in alveolar air after breathing oxygen for 7 min Abnormal lungs with localized areas of hypoventilation, however, do not "wash out" nitrogen with this rapidity and consequently have elevated alveolar concentrations at the end of this time

Low ventilation-perfusion ratios with increased venous admixture is the commonest cause of hypoxia in respiratory illness Bronchiectasis, pneumonia, cystic fibrosis, emphysema are some of the conditions in which this mechanism plays a fundamental role

Resistance to Diffusion of Oxygen from Alveolus to Capillary

The mechanism whereby oxygen passes from alveolus to pulmonary capillary was at one time a central issue preoccupying prominent departments of physiology at England's leading universities and prompting imaginative transcontinental expeditions to mountaintop laboratories (Henderson, 1938) J S Haldane at Oxford University contended that oxygen is secreted into the capillaries against a partial pressure gradient, whereas J Barcroft at Cambridge maintained that this transfer is entirely passive, following the laws of diffusion The issue has been resolved conclusively in favor of the latter, partly because methods for measuring gas pressures in blood improved It is now clear that venous blood is oxygenated because its P_{O_2} (about 40 mm Hg at rest) is lower than alveolar P_{O_2}, that the rate at which oxygen diffuses is a function of the partial pressure difference between alveolus and capillary, and that, under normal circumstances, virtually complete equilibrium is reached between alveolar and pulmonary capillary P_{O_2} (Fig 6 9) When resistance to diffusion is abnormally high, however, or when transit time through the pulmonary circuit is shortened, a measurable alveolar-capillary partial pressure difference may remain Under these circumstances, the total alveolar-arterial partial pressure difference is the sum of a diffusion gradient between alveolus and capillary plus the capillary to arterial difference attributable to anatomical and physiological venous admixture

Separating the diffusion from the venous admixture component of the total alveolar-arterial P_{O_2} difference is necessary for purposes of measuring the pulmonary diffusing capacity for oxygen Before passing to this topic, however, it should be clear that one of the causes of reduced P_{O_2} in arterial blood pathologically is incomplete equilibration of capillary blood with alveolar gas due to abnormal resistance to diffusion across the alveolar-capillary interface Interstitial fibrosis is the classic clinical condition producing hypoxia from this cause Increasing alveolar P_{O_2} by raising F_{I_O} is the logical way of relieving severe hypoxia from this cause

The Pulmonary Diffusing Capacity for Oxygen

The pulmonary diffusing capacity for oxygen (D_{O_2}) is the ratio of oxygen transfer in milliliters of oxygen per minute ($\dot{V}_{O_2}$) from alveolar gas to capillary blood divided by the pressure (mm Hg) required for this transfer

$$D_O = V_O / (P_{A_O} - \bar{P}_{C_O})$$

where P_{A_O} and $\bar{P}_{C_O}$ are partial pressures of oxygen in alveolar gas and pulmonary capillary, respectively (Forster, 1957) Inasmuch as the P_{C_O} continuously changes as blood flows through the capillaries, a mean value ($\bar{P}_{C_O}$) must be obtained as described in a subsequent paragraph

A normal D_{O_2} at rest is approximately 15 to 20 ml of oxygen/min /mm Hg P_{O_2} difference between alveolar and mean capillary oxygen tension The diffusing capacity reflects resistance to diffusion per unit area as well as total surface area for diffusion It is therefore reduced when the alveolar-capillary interface is thickened or where surface area for diffusion is attenuated It rises during exercise by as much as three or four times, presumably because the dimensions of the diffusing surface enlarge as additional capillaries open around each alveolus or as existing capillaries enlarge It increases from infancy to maturity and falls with age, probably reflecting changes in surface area for diffusion

Measuring D_{O_2} requires knowing V_{O_2}, or metabolic rate, P_{A_O} which can be calculated from the alveolar equation, and $\bar{P}_{C_O}$, which cannot be measured directly but which can be estimated from arterial P_{O_2} as follows· It was

indicated above that the total alveolar-arterial Po_2 difference is divisible into a diffusion and a venous admixture component The key to measuring $\bar{P}c_O$ is to create circumstances under which the venous admixture component vanishes or becomes negligible, for under this circumstance arterial Po_2 (Pa_O) approaches Pc_O, This can be done by producing hypoxia sufficient to displace Pa_O to the steep part of the dissociation curve, i e , the part for which substantial changes in saturation due to venous admixture result in relatively small changes in Pa_O To illustrate, it was shown above that the capillary-arterial Po_2 difference is 13 mm Hg if 6% of the cardiac output bypasses the lungs when PA_O and Pc_O are 100 mm Hg, and when the arterial venous saturation difference is 25% If this calculation is repeated at a saturation of 75% (Po_2 = 40 mm Hg), the saturation of venous blood will be 50% (Po_2 = 26 5 mm Hg), and the saturation of mixed arterial blood will be (0 06 × 50) + (0 94 × 75) = 3 00 + 70 50 = 73 5% The Po_2 at this saturation is 39 5 mm Hg, a difference of only 0 5 mm Hg from the pulmonary capillary Po_2 (40 mm Hg) By making the same calculation at a low Po_2, therefore, the venous admixture effect of 13 mm Hg was reduced to 0 5 mm Hg The reason for this striking difference is the characteristic of oxyhemoglobin saturation whereby a given change in saturation at high levels of saturation has a much greater effect on Po_2 than at low saturations

Measuring Do_2 requires not only eliminating the effect of venous admixture on the alveolar-arterial Po_2 difference but creating a measurable difference between PA_O, and Pc_O This can be accomplished by breathing low concentrations of oxygen and by exercise (Lillienthal et al , 1946) A low PA_O reduces the Po_2 gradient between alveolus and venous blood entering the capillary, the rate of diffusion is therefore slowed, the amount of oxygen taken up at any point along the capillary is less, and complete equilibration with alveolar gas is not achieved Exercise increases cardiac output and pulmonary blood flow, thereby reducing time for equilibration of Pc_O, with PA_O Both exercise and hypoxia, therefore, foster an end-capillary PA_O, − Pc_{O_2} sufficiently large to be measured reliably

In summary, by measuring the alveolar-arterial Po_2 difference at a high level of oxygenation to identify venous admixture, and at a low level of oxygenation during exercise, it is possible to create measurable differences between PA_O and Pa_O attributable to a difference between PA_O and Pc_O

BOHR INTEGRATION

The final step in calculating diffusing capacity is to derive a mean Pc_O ($\bar{P}c_O$) This can be done if one knows the Po_2 at the arterial and venous ends of the pulmonary capillary as well as at all intermediate points The former are measurable directly, the latter can be calculated from the oxyhemoglobin dissociation curve by Bohr integration (Fig 6 10) The fundamental premise is that distance or time (T) along the pulmonary capillary is proportional to the reciprocal of the tension difference between alveolar gas and capillary blood

$$[T \cong 1/(PA_O - Pc_O)]$$

This is true because the rate of diffusion, i e , the time (T) required for a given change in saturation (ΔSo_2) at a given point along a capillary, is proportional to the partial pressure difference between alveolus and blood at that point Thus,

$$\Delta So_2/T \cong (PA_O - Pc_O)$$

The data of Figure 6 10 were calculated as-

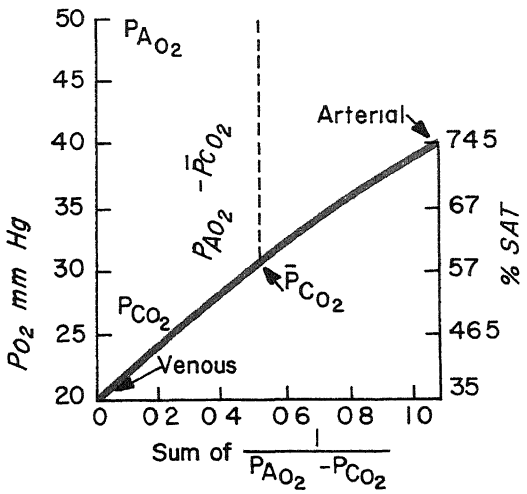


Fig 6 10 Change in P_O and oxyhemoglobin saturation during passage of blood through pulmonary capillary Data calculated by Bohr integration assuming conditions postulated in text Time or distance along capillary is a function of the sum of $1/PA_O - Pc_O$, as explained in text

suming the subject was breathing 14% oxygen, had an alveolar Po_2 of 50, an arterial saturation of 75% (Po_2 = 40 5) and an arterial-venous saturation difference of 40%, i e , mixed venous blood saturation of 35% and Po_2 of 20 mm Hg The end-gradient was 9 5 mm Hg and the mean gradient calculated by Bohr integration 29 mm Hg Assuming an oxygen consumption of 1740 ml /min , the diffusing capacity would be

$$1750 \text{ ml /min /29 mm Hg} = 60 \text{ ml /min /mm Hg},$$

a normal value during severe exercise

Although the circumstances of exertion and hypoxia required for this measurement are so Spartan as to limit its applicability, this method nevertheless remains the only direct approach to measuring the pulmonary diffusing capacity for oxygen Its ingenuity is obvious from the foregoing The premise that hypoxia itself does not affect what is measured may not be justified

Carbon Monoxide Diffusing Capacities

The use of carbon monoxide for measuring Do_2 is less trying on both subject and laboratory personnel None of the methods require hypoxia or exercise and some methods do not require analysis of blood samples

CO has a remarkable affinity for hemoglobin, its affinity being 210 times that of oxygen (Krogh, M , 1914) The Pco of plasma is therefore extremely low at a given concentration of HbCO Accordingly, plasma Pco can usually be assumed to be zero and the alveolar-arterial partial pressure difference becomes simply the alveolar Pco Analyzing blood and performing the equivalent of the tedious Bohr integration are therefore unnecessary

Dco can be converted to Do_2 by multiplying by 1 23 This figure indicates that oxygen is slightly more diffusible than CO This is because diffusibility is directly proportional to solubility and inversely proportional to the square root of the molecular weight Thus,

$$\frac{Do_2}{Dco} = \frac{Sol\ O_2}{Sol\ CO} \times \sqrt{\frac{m\ w\ CO}{m\ w\ O_2}} = \frac{0\ 0244}{0\ 0185} \times \sqrt{\frac{20}{32}} = 1\ 23$$

Diffusing capacity measured by whatever technique is subject to continuing scrutiny in efforts to define more precisely the significance of its measurement The complexity of the problem is apparent if one recalls that the lung is not one alveolus, one capillary, one dead space and one venous admixture as idealized in Figure 6 9, but millions of alveolar-capillary relationships covering a spectrum of ventilation perfusion-relationships Generally speaking, this spectrum is broader in disease than health (Bates and Christie, 1964), further complicating the interpretation of measurements designed to clarify the status of the lung

The concept of an alveolar-capillary "Membrane" is itself deceptive Oxygen and CO_2 diffuse through a number of layers in passing from air to hemoglobin or vice versa, including alveolar fluid, alveolar epithelium, tissue space between epithelium and capillary, capillary endothelium, plasma, red cell membrane and cytoplasm (De Reuck et al , 1962) There is evidence to suggest that the rate at which O_2 and CO combine with hemoglobin may limit Do_2 and Dco and that one therefore does not measure "membrane" properties so much as reaction rates One can conceive of O_2 transfer from alveolar air to red blood cell as diffusion across a series of resistances, for Do_2 has the dimensions of flow/pressure and $1/Do_2$, therefore, the dimensions of pressure/flow or resistance Roughton and Forster have subdivided total resistance into two resistances, one attributed to intracapillary processes (reaction rates with hemoglobin, diffusion in plasma and into red cell) and the other attributed to diffusion across the barriers separating air from plasma Thus,

$$1/D_L = (1/Dm) + (1/\phi Vc)$$

where D_L is diffusing capacity of the lung, Dm is diffusion across "membranes," Vc is the capillary blood volume and ϕ is the amount of oxygen taken up per unit time per milliliter blood of known hemoglobin concentration per milliliter Hg Po_2 Experimental data suggest that the intracapillary component ($1/\phi$ Vc) makes up a third or more of the total resistance One of the interesting applications of this formulation is its use for measuring pulmonary capillary blood volume (Vc). Normal values are on the order of 60 to 110 ml

Limitations and complexities not withstanding, however, it is clear that what diffusing capacity measures *1)* is reduced where histology of the lung suggests that the resistance to

diffusion is increased by thickening or change in the characteristics of the layers separating alveolar gas from pulmonary capillary blood, and 2) varies directly with changes in size of the pulmonary-capillary surface area for diffusion

Diffusion of CO_2

CO_2 is much more soluble in water than oxygen and its molecular weight is only slightly greater A comparison of its diffusibility with that of oxygen is as follows

$$\frac{D_{CO_2}}{D_{O_2}} = \frac{0\ 592}{0\ 0244} \times \sqrt{\frac{32}{44}} = 20\ 7$$

Because of this greater diffusibility, end-capillary alveolar P_{CO_2} gradients are generally thought not to occur

The dissociation curve for CO_2 is also very different from the oxyhemoglobin curve Whereas the latter is nonlinear, the CO_2 curve is relatively linear within physiological ranges of P_{CO_2} and CO_2 content It was explained above that it is because of the nonlinearity of oxyhemoglobin dissociation that local differences in $P_{A_{O_2}}$ result in an alveolar-arterial P_{O_2} difference Because CO_2 dissociation is more nearly linear, an alveolar-arterial P_{CO_2} difference from this cause is not likely to occur Thus, neither a diffusion gradient nor a venous admixture effect influence alveolar-arterial partial pressure relationships The validity of using arterial P_{CO_2} to measure alveolar P_{CO_2} rests on this statement

Hypoxia

This chapter includes discussion of four different causes of low arterial P_{O_2} for a subject breathing room air at sea level *1)* alveolar hypoventilation, 2) increased anatomic shunts, e g , right-to-left shunts in congenital heart disease, *3)* ventilation-perfusion abnormalities increasing venous admixture, *4)* increased resistance to diffusion The following characteristics are helpful in distinguishing among these physiologically when they are present in pure form Alveolar hypoventilation is invariably associated with elevated arterial and alveolar P_{CO_2}, the other causes are not Anatomic shunting is the only cause in which P_{A_O} remains relatively low when the patient breathes air with a high P_{O_2} The patient with hypoxia from a diffusion gradient commonly has a low alveolar P_{CO_2}, reflecting his attempt to raise alveolar P_{O_2} by increasing alveolar ventilation His arterial P_{O_2} drops precipitously with exercise because the increased requirements for oxygen diffusion can be met only by steepening the P_{O_2} gradient across the alveolar-capillary walls Moreover, his alveolar-arterial P_{O_2} difference remains high when he breathes an oxygen poor mixture This procedure, however, lowers the alveolar-arterial P_{O_2} difference when venous admixture is the primary cause of hypoxia

3 / THE MECHANICS OF BREATHING

The mechanics of breathing have captured the interest of workers in both the basic sciences and clinic, important contributions having been made by both groups individually or working together Fundamental concepts appeared in the German literature in the period 1915 to 1929, (Rohrer et al , 1925, Neergaard and Wirz, 1927) Technical advances in simultaneously recording volumes, flows, and small pressures with high fidelity have spurred rapid progress in this country and abroad, resulting in such feats as measuring the forces required for inflating the lung during the first breath of a newly born infant Most prominently associated with this activity in this country have been groups originally associated with the Departments of Physiology at the University of Rochester, at the School of Public Health of Harvard University, and at the Graduate School of the University of Pennsylvania Excellent reviews are included in four texts dealing with the respiratory system listed in the bibliography (Avery, 1964, Bates and Christie, 1964, Chernıak and Chernıak, 1962, Comroe and Mortimer, 1964)

Muscles of Respiration

Supplying oxygen to the alveoli and removing carbon dioxide from the alveoli is accomplished by the flow of air in and out of the lungs with each breath This is a mechanical process for which the muscles of respiration provide the force

DIAPHRAGM

Quiet breathing results from alternate contraction and relaxation of the diaphragm and intercostal muscles The former is attached circumferentially to fixed structures, vertebrae and sternum, and to relatively mobile structures, the lower ribs It is normally dome-shaped (Fig 6 11), its fibers run nearly vertically upward near its periphery and horizontally at its tendinous central portion Contraction during inspiration produces increased tension within the diaphragm itself and on its attachments Its central tendon is therefore pulled down, increasing the vertical dimensions of the thorax, while the lower ribs are moved somewhat upward, increasing the lateral and anteroposterior dimensions of the lower chest

The effectiveness of the diaphragm for changing the dimensions of the chest is related to the strength of its contraction and to its contour when relaxed Strength of contraction, as in any muscle, is a function of the number of motor units activated and of the synchronization and frequency of their discharge Interruption of both phrenic nerves immobilizes the diaphragm without changing its configuration Increasing the circumference of the lower chest pathologically, as in obstructive pulmonary disease, flattens the diaphragm, consequently immobilizing it to varying degrees irrespective of how forcefully it contracts

Descent of the diaphragm can be accomplished only by displacement of abdominal contents This is made possible by relaxation of the muscles of the abdomen during inspiration Although abdominal pressure does not normally interfere with the diaphragm's mobility, advanced pregnancy, extreme obesity, excessive flatulence, tight abdominal garments, and other circumstances can

The diaphragm probably contributes more to inspiration than do the intercostals This statement rests on estimates of vertical displacement (about 1 5 cm) and area (about 350 cm ^{2}) of the diaphragm during quiet breathing Since the volume of a single, quiet inspiration (tidal volume) is 500 to 700 ml , these estimates suggest that normal breathing is predominantly diaphragmatic The relative proportions between diaphragmatic and intercostal contribution probably vary among different individuals and with different age groups The ribs of infants, for example, are normally virtually horizontal Inspiration must therefore result almost exclusively from increase in the vertical dimensions of the chest by contraction of the diaphragm

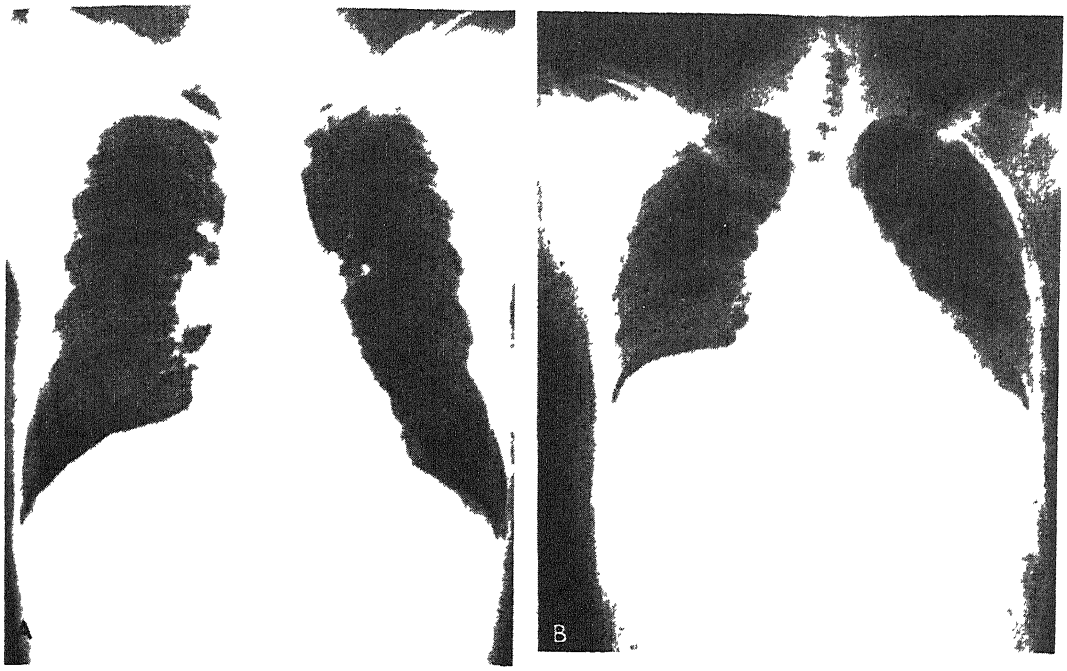


Fig 6 11 Roentgenograms of the chest taken in the antero-posterior projection Film on the left was taken after complete inspiration when lungs contain the total lung capacity Film on the right was taken after forced expiration when lungs contain the residual volume Note that the change in vertical dimensions of the chest is considerably greater than the change in transverse diameter

INTERCOSTAL MUSCLES

The action of the diaphragm is straightforward and clearly inspiratory The action of the muscles running between the ribs, however, is somewhat more complicated The external intercostals run forward and downward whereas the internal intercostals run backward and downward, i e , approximately at right angles to the external Both groups originate and insert on adjacent ribs and can be expected to bring these ribs closer together when contracting Whether a given rib moves upward, downward, or remains stationary, depends on the relative stability or fixation of adjacent ribs

Inspiration is associated with electromyographic evidence of intercostal activity (Bronk and Ferguson, 1935) The cephalad motion of the ribs indicates that the upper ribs are fixed relative to the lower When contraction of the abdominal muscles fixes the lower ribs during cough or forced expiration, the rib cage moves down, and electromyography again shows intercostal activity The intercostals contract during quiet inspiration, forced inspiration, and forced expiration, but not during quiet expiration (Campbell, 1958) Whether the internal and external groups have distinct functions in man is not clearly established

Adequate breathing continues in the absence of either the diaphragm or the intercostals In the absence of both, however, exogenous sources of power are usually needed for maintaining sufficient ventilation for life Inasmuch as the intercostals and diaphragm are innervated from separate segments of the spinal cord (C3 and 4 for the phrenic nerves as compared to thoracic segments for intercostal nerves), paralysis of one is not necessarily associated with paralysis of the other

MOTION OF THE RIBS

The ribs are more or less semicircular bones articulating with the vertebrae at one end and with the sternum or costal cartilages at the other Rib motion, therefore, includes rotation around an axis joining these two points of attachment ("bucket-handle motion") (Fig 6 12) In addition, the motion of the ribs includes elevation and depression of the sternal end Because this sternal end is lower than the vertebral end, elevating the shaft of the rib causes the sternum to protrude, increasing the antero-

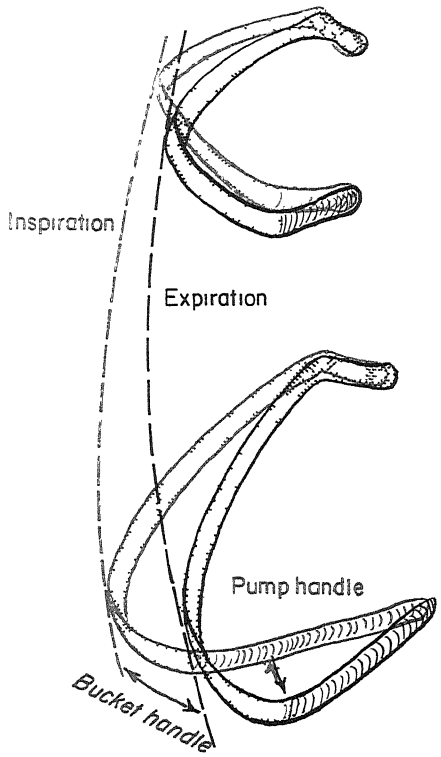


Fig 6 12 Motion of upper and lower ribs Solid outlines indicate position during inspiration (From Cherniak and Cherniak 1961)

posterior diameter of the chest Depression of the shaft of the rib correspondingly causes the opposite Inspiration, therefore, is associated with an increase in the vertical diameter of the chest due to contraction of the diaphragm, with an increase in transverse diameter due to the "bucket-handle" action of the ribs, and with an increase in anteroposterior diameter due to the effect of rib motion on the sternum

The motion of the ribs is not equivalent throughout the rib cage (Fig 6 12) The upper ribs are arcs of smaller circles and are more nearly horizontal than the lower ribs Moreover, the lower ribs on inspiration are affected by displacement of all superior ribs, their motion is, therefore, the sum of the narrowing of many interspaces rather than only the adjacent one Because of these differences, the changes in lateral diameter of the chest are greater in the lower than in the upper thorax (Fig 6 11)

As in the instance of the diaphragm, the contribution of costal motion to breathing depends on the force of intercostal contraction and on the position of the ribs prior to intercostal contraction. Accordingly, abnormal flaring or elevation of ribs and sternum interfere with the costal contribution to inspiration

ACCESSORY MUSCLES

Electromyographic evidence from man and action potential recordings from the phrenic and intercostal nerves of animals (Adrian and Bronk, 1929) indicate that *1)* quiet inspiration results from contraction of diaphragm and intercostals, and *2)* quiet expiration occurs in the absence of any muscular contraction whatsoever The accessory muscles are normally brought into play only for forced inspiration and expiration Many such muscles could be listed The distance runner struggling for air, for example, may use even the platysma for expanding his chest and the patient in paroxysms of cough probably contracts every muscle of trunk, thorax, and pectoral girdle during forced expiration The principle accessory muscles, however, are the anterior neck muscles (sternocleidomastoids and scalenes), which are used during forced inspiration to stabilize and raise the first ribs and upper sternum, and the abdominal muscles, which are essential for forced expiration The latter, together with the diaphragm, are used also for raising intra-abdominal pressure during defecation, urination, vomiting, parturition, cough, and the Valsalva maneuver The patient with paralysis of abdominal muscles is therefore at a considerable disadvantage

Lung Volumes

The lungs and chest return passively to their original volume after inspiration This end-expiratory volume is known as the functional residual capacity It is the volume which reflects the equilibrium position of the chest and lungs when undisturbed by muscle contraction It is also the volume into which each tidal volume is diluted during inspiration

By adding the inspiratory capacity to this volume during a maximal inspiratory effort, the lung volume is increased by about 2½ and contains the total lung capacity On the other hand, delivering the expiratory reserve volume by forced expiration reduces the end-expiratory volume by about one-half to the residual volume The vital capacity is the volume which can be expired after a maximal inspiration by forceful expiration, i e , it is the sum of the inspiratory capacity and the expiratory reserve These relationships are shown in Fig 6 13

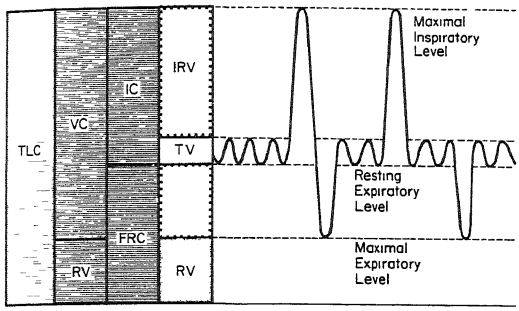


Fig 6 13 The subdivisions of the total lung capacity *TLC* = total lung capacity *VC* = vital capacity *RV* = residual volume *IC* = inspiratory capacity *FRC* = functional residual capacity *IRV* = inspiratory reserve volume *TV* = tidal volume *ERV* = expiratory reserve volume (From Comroe et al 1962)

The vital capacity and its subdivisions can be measured by collecting and measuring, in a spirometer, the volume of gas which flows into or out of the lungs during normal or forced inspiration and expiration However, the residual volume, functional residual capacity, and total lung capacity cannot be directly recorded These are commonly measured by "nitrogen wash-out" or "helium dilution " In "nitrogen wash-out" the subject is made to breathe nitrogen-free oxygen long enough (about seven minutes) to fully "wash-out" the nitrogen from his lungs His expired air is collected, its volume is measured, and it is analyzed for concentration of nitrogen The total amount of nitrogen collected is the product of the concentration and volume Since the original concentration of nitrogen in the lungs is known (about 80%) the lung volume can be determined Thus, if 50,000 ml of expired air were collected over the 7-min period and the concentration of nitrogen in the collected gas was 4%, the functional residual capacity would be 50,000 × 0 04 − 0 80 or 2500 Helium dilution is comparable in concept A measurement is made of the dilution of helium which results from adding the volume of the lungs to a closed circuit containing a known amount of helium

Whatever method is used, the functional residual capacity can be measured more reproducibly than the residual volume or total lung capacity because it is independent of muscular effort The residual volume is usually calculated by subtracting the expiratory reserve from the functional residual capacity, the total lung capacity by adding the inspiratory capacity

There are numerous normative data on lung volumes One commonly used empiric prediction formula for adults is (Baldwin et al , 1948)

Males, vital capacity = [27 63 − (0 112 × age)] × height in centimeters,
Females, vital capacity = [21 78 − (0 101 × age)] × height in centimeters

These reflect the fact that vital capacity is related to age, sex, and height A normal person's total lung capacity and residual volume can be estimated by knowing that the vital capacity comprises about 70% or 80% of the total lung capacity

The vital capacity and total lung capacity are reduced by weakness of the normal and accessory muscles of inspiration and expiration, by immobilization of the ribs and sternum, or by space-occupying defects of the chest The vital capacity is less in the supine position than in the upright position because the intrathoracic blood volume is increased in the supine position, displacing an equivalent volume of air

The functional residual capacity is the volume to which the lungs and chest arrive in the absence of muscle contraction It represents, therefore, an equilibrium position of the chest and lungs, intermittently disturbed by the forces of muscle contraction Factors affecting this equilibrium and its disturbance are discussed below

Pressures Required for Changes in Volume

THE LUNG

The chest wall, independent of the lung, and the lung itself are both hollow structures with elastic properties Each has a certain volume when pressures inside and outside are equal and each responds to a distending or retracting force by a change in volume proportional to that force The relationship between volume change and pressure required to inflate the lung alone is shown by the dotted line of Fig 6 14 It shows that the equilibrium volume of the lung, when no inflationary pressure is applied, is a small proportion of the total lung capacity (approaching 0 in the figure) and that increments in inflationary pressure produce specific increments of volume The slope of this line indicates the compliance of the lung, i e , the relationship between volume and pressure Compliance is equal to the volume change in liters divided by pressure change in centimeters of H_2O Compliance decreases with inflation,

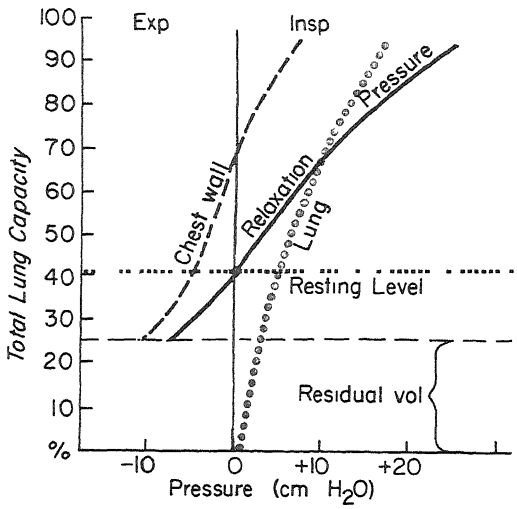


Fig 6 14 Pressure-volume relationships of chest and lung in relaxed subject (relaxation pressure) in lung alone and in chest wall alone See text for explanation (Adapted from Rahn et al 1946)

indicating that as the lung inflates, progressively greater force is required for a given volume change Under normal circumstances in man, it is not possible to measure the pulmonary compliance at volumes below the residual volume, for this would require opening the chest Compliance is therefore usually measured over the relatively linear segment of the curve which includes the volumes of the lung during normal breathing, i e , the functional residual capacity plus the tidal volume or more Normal values in man are approximately 0 2 liters/cm of H_2O (Frank et al , 1956)

The more a given lung is inflated the less is the compliance It is less for small lungs than for large lungs (0 0012 for rats as compared to 0 09 for dogs) and is reduced in man by circumstances which reduce the amount of lung, such as surgical removal of lung tissue It is also reduced by change in the composition of lung tissue (pulmonary fibrosis) and progressively, with time, if the distension of the lung with each breath is unvarying (Ferris and Pollard, 1960)

The fact that compliance is less for small lungs than for large lungs (Crosfill and Widdicombe, 1961) and less for a highly inflated lung than for less inflated lungs does not necessarily indicate fundamental differences in stiffness True difference in elastic properties between one lung and another are reflected only when correction is made for differences in size or volume. One approach is to divide compliance by the functional residual capacity (so-called "specific" compliance), normal values in adults are between 0 05 and 0 06 liters/ml of H_2O per liter Another approach is to measure the increase in transpulmonary pressure required to inflate the lungs to a volume midway between any resting level and full inspiration, normal values in man are 7 58 ± 2 53 cm of H_2O (Salazer and Knowles, 1964)

THE CHEST

The pressure-volume curve of the chest wall in the absence of the lung is displaced far to the left of the lung curve (Fig 6 14) This reflects the fact that the volume of the chest at a given distending pressure is considerably greater than the volume of the lung at that pressure Thus, although the volume of the lung when no pressure is applied is small, the volume of the chest is about 70% of the total lung capacity It is necessary to apply negative pressure to the chest (inside negative with respect to outside) in order to reduce its volume to the functional residual capacity, positive pressure, on the other hand, (inside positive with respect to outside) is required for inflating the lung to the functional residual capacity The compliance of the chest wall is about 0 2 liters/cm of H_2O

THE LUNG AND CHEST

Up to this point, the lung and chest have been treated as separate structures If the lung is now placed within the chest and its volume is to be the functional residual capacity, it will recoil and pull in on the chest with a force equal to the pressure required to distend it to the functional residual capacity (about 5 cm of H_2O) This retracting force of the lung on the chest wall is the basis of negative (below atmospheric) pressure between the visceral and parietal pleurae (interpleural[1] "space") As a result of it, the chest volume is less than it would be in the absence of the lung by a volume equal to the difference between the functional residual capacity and 70% of the total lung capacity

The relaxation pressure curve of an intact subject (Fig 6.14) depicts the relationship between pressure of air in the lungs and the

[1] "*Inter*pleural" is used here and subsequently to indicate between the visceral and parietal pleural surfaces "*Intra*pleural" is more commonly used, referring to the pleural cavity Inasmuch as there is normally no cavity or space as such, interpleural seems more appropriate

volume of the lung-chest system when the subject is completely relaxed The pressure is 0 at the functional residual capacity, for the elastic recoil of the lung is matched by an equal and opposite pull by the tensile properties of the relaxed chest wall Increasing the volume of the chest and lung above the functional residual capacity requires increasing the air pressure This pressure, or force, is equal to the force generated by contraction of the muscles of inspiration Decreasing the volume of the chest and lung below the functional residual capacity requires reducing the intrapulmonary pressure below atmospheric pressure, this pressure is a measure of the force generated by the muscles of forced expiration Fig 6 14 shows that the force developed by the muscles of inspiration up to a volume comprising about 70% of the total lung capacity is expended only on the lungs At higher volumes, however, force is needed for expanding the lungs as well as the chest During forced expiration, the energy of muscle contraction is expended exclusively on the chest wall The compliance of the chest and lung (slope of the relaxation pressure curve) is about 0 1 liter/cm of H_2O, i e , one-half the compliance of lung or chest individually

It is apparent that the relationship between lung and chest can be compared to that between two opposing coiled springs whose equilibrium is disturbed in one direction or the other by a transient force equivalent to contraction of the muscles of respiration The pressure which represents the recoil of the lung away from the chest wall (or the force required to keep the lung inflated at a given volume) can be measured in the intact chest by carefully placing a needle between the visceral and parietal pleurae, injecting a small bubble of air, and recording the pressure in the bubble with a manometer It can be measured more conveniently and safely by measuring the pressure in a structure within the chest but outside the lungs The esophagus is the most commonly used of these, pressure within it being recorded by having the subject swallow a small compliant balloon partially filled with air and recording the pressure within the balloon (Buytendijk, 1949, Mead, 1961)

LUNG CHEST, AND MUSCLE CONTRACTION

The maximum pressures which can be developed within the lungs by forcefully contracting the muscles of inspiration or expiration are

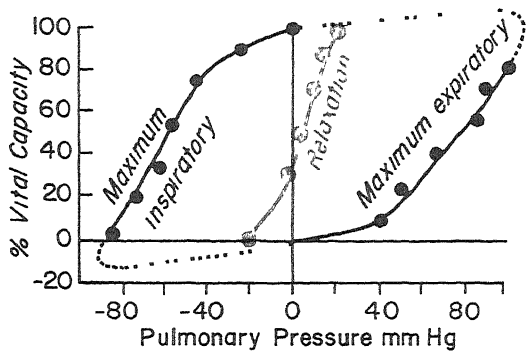


Fig 6 15 Pressure volume diagram of intact chest including the relaxation pressure curve (From Fenn 1951)

shown in Fig 6 15 The expiratory pressures are greater as lung volume increases, partly because the muscles of forceful expiration are stretched to a length at which they can develop greater tension, and partly because the recoil of the lung-chest system toward a smaller volume is greatest when the chest and lung are fully expanded The inspiratory pressures, represented by negative intrapulmonary pressure, are greater at low lung volumes because the muscles of inspiration are at a favorable portion of their length-tension relationship and because recoil of the chest toward a more inflated state is maximal The relative portion of the total pressure attributable to the muscle contraction and to recoil at each volume is represented by the difference between the relaxation pressure (recoil) and maximum pressure (recoil plus muscle contraction)

Ventilation

The volumes and pressures discussed in the previous section are independent of air flow or time Ventilation has the dimensions of both volume and time Normal ventilation in resting day-old infants (2 5 kg) is about 500 ml /min , with a respiratory rate of about 33 breaths/min , and a tidal volume of 15 ml (Smith, 1959) By adulthood (70 kg), normal ventilation has increased to about 6000 ml /min , at a respiratory rate of 12 breaths/min and a tidal volume of 500 ml Thus, the infant ventilates about 200 ml /min /kg as compared to less than one-half this amount in an adult

The flow of air into and out of the lung varies from instant to instant during normal breathing (Fig 6.16) If one inspires maximally to the total lung capacity and then expires as forcefully as possible, the plot of volume and

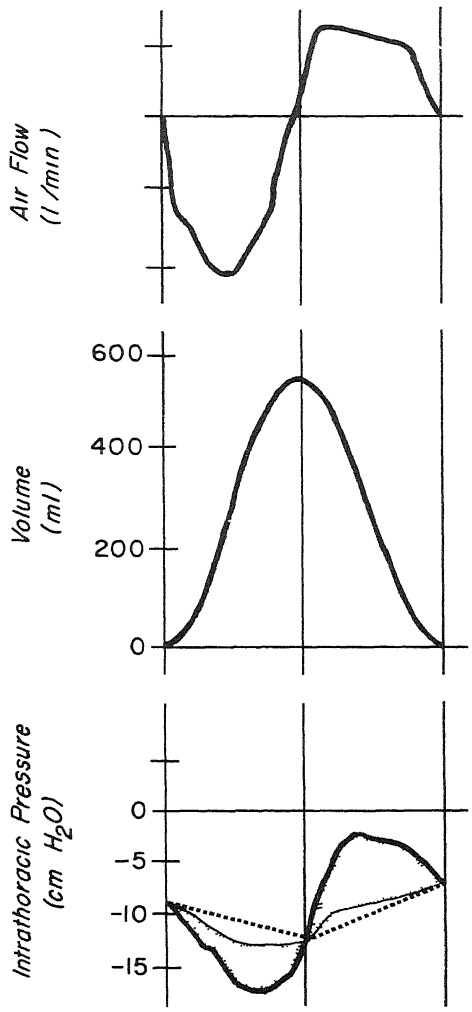


Fig 6 16 One breathing cycle showing air flow one tidal volume and intrathoracic pressures on simultaneous ordinates Inspiration is indicated by downward deflection on flow tracing and upward deflection on volume record See text for subdivision of intrathoracic pressure record (From Cherniak and Cherniak Saunders Philadelphia 1961)

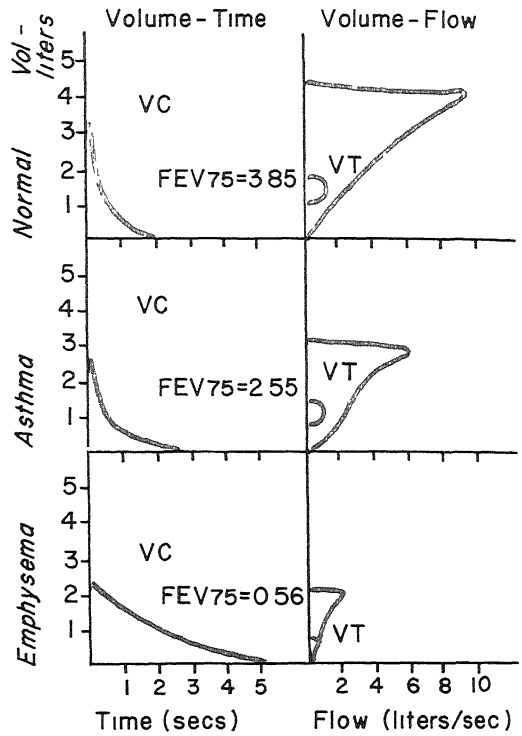


Fig 6 17 Volume of air forcefully exhaled after maximal inspiration Abscissa is time on left half of figure and flow on right half VC = vital capacity VT = tidal volume FEV 75 = forced expiratory volume in 75 seconds 1) Flow normally slows progressively as volume decreases 2) peak flows attainable in asthma and emphysema are low 3) flows during tidal ventilation are a small fraction of peak flows in normal individuals but a large fraction in emphysema 4) because the functional residual capacity is so elevated the patient with emphysema has a much reduced reserve of volume to which to inflate his lungs for attempting to increase flow (From Bates and Christie Saunders Philadelphia 1964)

flow rate as functions of time appears as in Figure 6 17 (normal) One can calculate from such tracings that the maximum expiratory flow rate exceeds 400 liters/min and the proportion of the vital capacity exhaled in 1 sec, 2 sec, and 3 sec exceeds 83%, 92%, and 97% respectively These values are all used in comparing normal and abnormal patterns of ventilation They are low where airway diameter is reduced, asthma and emphysema being the classic examples of obstructive respiratory abnormality (Fig 6 17) (Daymen, 1951)

PRESSURES REQUIRED FOR AIR FLOW

Ventilation, or air flow in and out of the lungs, occurs as a result of a pressure difference between the alveolar and oral ends of the airways During normal inspiration, the pressure of air in the alveoli falls transiently below atmospheric pressure It equals atmospheric pressure at the end of inspiration and rises above atmospheric pressure during expiration The velocity of air flow is proportional to the pressure difference The relationship between pressure difference and flow depends on the type of flow Two types or patterns are commonly recognized, laminar flow and turbulent flow The equation for laminar flow is expressed in Poiseuille's law, in which the pressure (P) differ-

ence between the two ends of a tube through which there is flow is directly proportional to the volume flowing per unit time (Q), to the viscosity of the fluid (v), to the length of the tube (L) and inversely related to the fourth power of the radius (r)

$$P_1 - P_2 = Q8LV/2\pi r^4$$

It is apparent from this equation that pressure is extremely sensitive to radius, a 2-fold decrease in radius of an airway, for example, requires a 16-fold increase in pressure if flow is to remain constant It is also apparent that pressure is directly and simply related to flow, e g , doubling pressure doubles flow ($P_1 - P_2 = KQ$) in which K is a constant incorporating the other terms

As flow increases in a tube of given dimension, a rate is reached at which the flow pattern ceases to be laminar and becomes turbulent Turbulence occurs when the product of density (d), linear velocity (V), and diameter (D) divided by viscosity (v) exceeds 2000

$$\text{Reynolds number} = dVD/v$$

The relationship between pressure and flow in the presence of turbulence is exponential, pressure being related not simply to the velocity of flow but to the square of velocity, e g , doubling flow requires a fourfold increase in pressure

$$P_1 - P_2 = KQ^2$$

Calculations of Reynolds numbers in different parts of the airways suggest that flow is predominantly laminar during quiet breathing but predominantly turbulent during even moderate hyperventilation Inasmuch as turbulence is favored not only by the factors included in calculating the Reynolds number but also by irregularities and branching of tubes, it is reasonable to assume that even during quiet breathing the pressures for moving air are required for both laminar and turbulent flows It is estimated that about 10% of flow resistance is due to turbulence during quiet breathing

RESISTANCE TO AIR FLOW

$$\text{Resistance} = \frac{\text{pressure}}{\text{flow}} = \frac{\text{cm } H_2O}{\text{liters/sec}}$$

Flow resistive pressures of the airways can be measured if flow and pressure difference between alveoli and environment are known Regional resistances at intermediate points between the alveoli and the room can also be determined if pressure is known at appropriate sites By measuring pressures in the pharynx with a small balloon, in the trachea by inserting a needle between the first and second tracheal rings, and in the alveoli by an indirect method described below, oral or nasal resistances, pharyngeal-laryngeal resistances, and intrapulmonary airway resistances can all be calculated (Ferris et al , 1964)

Alveolar pressures are measured with a body plethysmograph (Du Bois et al , 1956) A subject is enclosed in an airtight chamber and breathes into and out of this chamber The air within the chamber will be compressed and its pressure will rise whenever the external dimensions of the chest increase more rapidly than an equivalent volume of air is taken from the chamber into the lungs This occurs during inspiration when the fall in intrapulmonary pressure required for air flow is associated with rarefaction of gases within the lung until flow of air from the chamber equalizes pressures in lung and chamber at a new lung volume During expiration, the dimensions of the chest decrease more rapidly than the tidal volume is exhaled, the chamber pressure consequently falls as intrapulmonary pressure rises until pressure equilibration is again restored at the end of expiration The plethysmographic pressure, therefore, reflects intrapulmonary pressure, the precise relationship being defined by calibration An equivalent experiment can be conducted recording volume change in the plethysmograph at constant pressure rather than pressure at constant volume (Mead, 1960, Jaeger and Otis, 1964) Resistance can be calculated if instantaneous flow is recorded simultaneously with pressure

Normal values (centimeters of H_2O/liter/second) for total airway resistance during panting are 1 50 ± 0 49 cm of H_2O/liter/sec Normal values during quiet breathing are 1 26 ± 0 39 Resistances for women are about 20% higher than for men, presumably because the airways of women are narrower

About 20% of the oral-alveolar air flow resistance can be attributed to flow through the mouth, 27% to flow through the glottis and larynx, and about 53% to flow through the airways below the second tracheal cartilage During quiet breathing through the nose, on the other

hand, the nasal passages contribute about two-thirds of the total air flow resistance Resistance offered by the larynx and glottis during normal expiration is about twice its value in inspiration Resistances are least, therefore, during inspiration, using oral rather than nasal breathing and in men rather than women (Ferris et al , 1964)

RESISTANCE OF LUNG (VISCOUS RESISTANCE)

Alveolar pressures during ventilation are reflected in interpleural pressures as we have seen In addition, however, interpleural pressures register the force required for overcoming the viscous resistance of the lung (Marshall and Du Bois, 1946, McElroy et al , 1955) This latter has been estimated variously to comprise a negligible to substantial (28°C, 21°C, and 30°C to 40°C) proportion of the total pulmonary resistance, i e , air flow resistance plus tissue viscous resistance Although this figure, as many others concerning respiratory mechanics, are educated approximations, it is clear that interpleural pressure during ventilation represents *1*) compliant characteristics of the lung, *2*) air flow resistances through the respiratory tract, and *3*) tissue resistance of the lung

The components of the total pressure difference between atmosphere and esophagus which are attributable to elastic (compliant) characteristics and to flow-resistance are represented in Figure 6 16 (intrathoracic pressure) The elastic component is indicated by the thin solid line It is not a straight line (cf dotted line) because the relationship between lung volume and time is not linear The shaded portion is the additional pressure required for air flow and tissue components It is greater during expiration than during inspiration because air flow resistance is higher.

RESISTANCE OF CHEST (VISCOUS RESISTANCE)

Resistance of tissues of the chest have also been estimated. From 30 to 40% of the total respiratory resistance is attributed to this source both during quiet oral breathing (0 4 liter/sec) and during mild hyperventilation (flows of 1 liter/sec) (Fig 6 18)

INERTIA OF AIR, LUNG, AND CHEST

The final component of the total mechanical force for respiration, inertia, has been measured for lung and air (Mead, 1956) At atmospheric pressure, a force of only approximately 0 01 cm of water is required to produce a volume acceleration of 1 liter/sec /sec Clearly, this is a small component of the forces required for ventilation

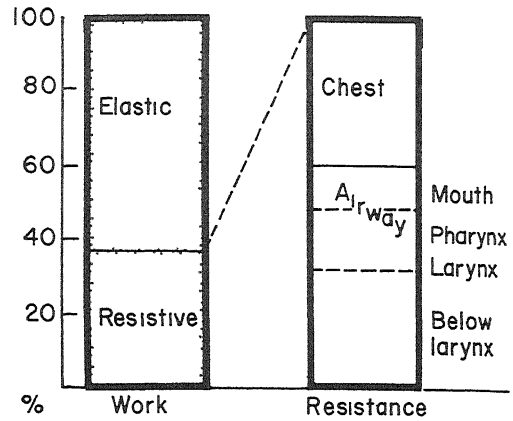


Fig 6 18 Subdivision of respiratory work and resistance Values for work bar for subject breathing 15 breaths per minute at a tidal volume of 500 ml Values for resistance bar for subject breathing through mouth Viscous resistance of lung was assumed to be 0 (See text)

The Work of Breathing

The chest and lungs are a pump operating against resistances to generate the pressures outlined above The concept of work is useful in expressing the energy required or expended for breathing (Otis, 1954)

$$\text{Work} = \text{force} \times \text{distance}$$

Force is equivalent to pressure (cm of water), and volume (ml) to distance, i e , ventilating 1 liter of air with a force of 1 cm of water pressure is equivalent to moving 1 ml of water (1 g) 1000 cm or doing 1 g -m of work

The total work of breathing has been estimated by placing a relaxed subject in a respirator and continuously measuring the pressure gradient between mouth and outside of the patient's body simultaneously with volume At a tidal volume of 500 ml and a frequency of 15 breaths per min , the work on the lungs, chest, abdomen, and air was approximately 0 32 kg -m /min The work done to stretch chest and lungs, i e , elastic work, was about 63% of the total work (Fig 6 18)

The work of breathing can be conceptualized by graphic representation The work done on the lungs during a single breath is shown in Figure 6 19 Oral-esophageal pressure differ-

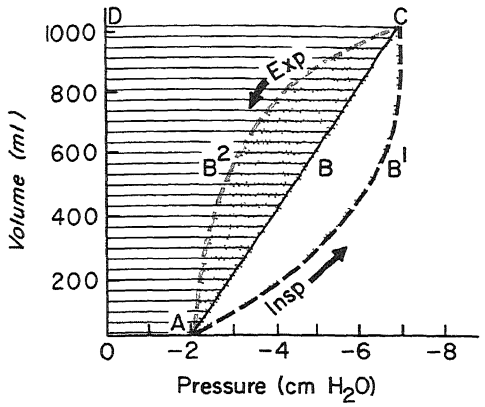


Fig 6 19 The relationship between lung volume and transpulmonary pressure during one inspiration and expiration of 1 000 ml See text for explanation (From Cherniak and Cherniak Saunders Philadelphia 1961)

ence is plotted on the horizontal axis and volume at any pressure on the vertical axis Point *A* is the plot of pleural pressure at the end of expiration when flow is 0, and volume is the functional residual capacity Point *C* plots pleural pressure at the end of inspiration when flow is 0 again, and volume is the functional residual capacity plus 1000 ml The line *AC* is, therefore, the compliance of the lung It is assumed to be straight because compliance is relatively linear over this range of lung volumes The elastic work (pressure × volume) done on the lungs is represented by the area of the trapezoid *ABCDO*

The pressures actually recorded during a breathing cycle are indicated by the loop AB^1CB^2 The area AB^1CB represents the nonelastic, i e , flow resistive, work done on the lungs and air during inspiration The area CB^2AB represents the nonelastic work done on lungs and air during passive expiration This is a portion of the elastic energy stored in the lungs at the end of inspiration The balance of the elastic energy, AB^2CDO, is not recovered during expiration and in this sense is "lost," possibly as heat

The shape of the loop AB^1CB^2 reflects nonelastic pressures It is, therefore, broad or more circular when flow rate is high (hyperventilation) or resistance increased (airway narrowing) Because the force required to stretch the lungs (elastic properties) is independent of the rate at which the lung is inflated, progressively increasing rates of ventilation augment flow-resistive work without affecting elastic work. Increased depth of breathing, on the other hand, increases elastic work, particularly since compliance decreases at higher lung volumes

Breathing is most economical when the balance of elastic and resistive work yields the lowest sum compatible with ventilation sufficient to meet the body's metabolic requirements It might be assumed that this would result in low tidal volumes and rapid breathing rates, for this type of breathing would sharply reduce elastic work However, as was brought out in the previous chapter, carbon dioxide excretion and oxygen supply are a function not of total ventilation but of alveolar ventilation

A tidal volume equal to the dead space (150 ml), for example, might accomplish a normal total ventilation (6 liters/min) at minimal elastic work but would be entirely wasted because no carbon dioxide would be excreted and alveolar oxygen pressure would not be maintained It is one of the important facts of biological economy that quiet respiration is accomplished at a tidal volume and frequency such that the sum of elastic and flow resistive work at an adequate alveolar ventilation is minimal This has been shown for man as well as for other animals (Otis, 1954)

The work of breathing can be expressed in terms of oxygen utilization This is about 0 5 ml of oxygen/liter of ventilation in normal individuals At a normal ventilation of 6 liters/min this is equivalent to 3 ml /min or about 1% of normal total requirements (300 ml /min) The oxygen cost increases with ventilation in a nonlinear manner, indicating progressively rising cost with the hyperventilation of exercise One estimate of oxygen utilization for breathing during severe exercise at a ventilation of 160 liters/min is 406 ml /min

Patients with obstructive disease expend a considerably larger proportion of their basal energy for breathing, for both resistance and ventilation are increased The augmented oxygen utilization and CO_2 production resulting from greater activity of respiratory muscles may place a significant additional burden on a marginal system

EFFICIENCY

The efficiency of the muscles of respiration can be calculated from the work they perform and the oxygen cost of doing this work

Efficiency = mechanical work/oxygen used

Values for mechanical work are about 0 03 kg-m /liter as compared to 0 6 kg -m /liter for the

work equivalent of oxygen utilization The efficiency, therefore, is about 5% This low mechanical efficiency indicates that most of the energy for respiration is in the form of heat

INTERPLEURAL PRESSURES

The pressure within the chest but outside the lungs measures both the elastic recoil of the lungs and the flow-resistance of air and lungs It is, therefore, most negative at high volumes, high inspiratory flow rates, and high inspiratory resistances It is subatmospheric throughout inspiration and expiration (Fig 6 16) except when the energy stored in the lungs at the end of inspiration is inadequate for the work of expiration This occurs during forced expiration, cough, sneeze, and expiration through narrowed airways In these instances, active work is done during expiration which must be added to the work of inspiration for calculating the total work of breathing It would appear graphically as an area extending to the left of the vertical axis in Figure 6 19

The fact that interpleural pressure is subatmospheric has important implications There will be flow into the thorax from any site at or above atmospheric pressure provided a flow-way is present Accordingly, air flows into the chest from the atmosphere and creates an interpleural space (pneumothorax) when the chest wall is perforated, venous blood flows toward the thorax and right auricle down the pressure gradient from the abdomen or other body part to the chest This gradient with respect to the abdomen is increased during inspiration not only because intrathoracic pressure falls, but because the descent of the diaphragms raises abdominal pressure by a few centimeters of water above atmospheric pressure Reducing or reversing this gradient by performing the Valsalva maneuver can compromise venous return sufficiently to cause hypotension and syncope

Another important indirect consequence of negative interpleural pressure concerns its effect on the patency of airways To the extent that these are exposed to interpleural pressures, they are subjected to a transmural force This force is a distending force at high lung volumes and rapid inspiration, a weaker distending force during passive expiration as interpleural pressure falls, and a compressing force during forceful or high resistance expiration when interpleural pressure rises above atmospheric pressure.

The pressure gradient across the airway wall depends not only on interpleural pressure but also on the pressure within the airway This is atmospheric when there is no air flow and the glottis is open During ventilation, it is more nearly atmospheric in the larger airways closer to the atmosphere and most nearly interpleural in the finer airways farther downstream on inspiration and farther upstream on expiration Accordingly, the airways subjected to the greatest transmural pressures are the larger ones Providing these with supporting cartilage is, therefore, good teleology Despite their cartilaginous support, however, these can be shown roentgenographically to undergo visible narrowing during forced expiration—a change which increases flow resistance

Airway patency is preserved not only by negative interpleural pressure and by cartilage, but by radial traction from the elastic tissue within which the airways are suspended This traction is a function of the relationship between length and tension of the lungs' elastic fabric, the number of elastic strands pulling on the airway per unit area of surface, and the degree to which the lung is stretched (lung volume) The significance of this support is shown in normal breathing by progressive slowing of air flow during forced expiration (Fig 6 17), indicating narrowing of the airways as lung volume decreases This narrowing is particularly pronounced in emphysema (Fig 6 17) where loss of elastic properties of lung tissue and the thinning out of parenchymal strands surrounding airways predisposes them to collapse especially during expiration This may be so pronounced that air flow ceases during expiration at lower volumes no matter the force with which the muscles of expiration contract A vicious cycle results as expiratory effort increases, interpleural pressure rises, airways undergo additional compression, and airflow is further slowed Breathing at pathologically high lung volumes, one of the hallmarks of obstructive abnormality, is a logical recourse

HYSTERESIS AND SURFACE ACTIVITY

The change in volume of the lung with pressure has been described during inspiration (Fig 6 14) During expiration, when the lung is allowed to deflate, the curve of its volume-pressure behavior does not retrace itself, but is displaced to the left forming a loop (Fig 6 20) The area of this loop represents an amount of

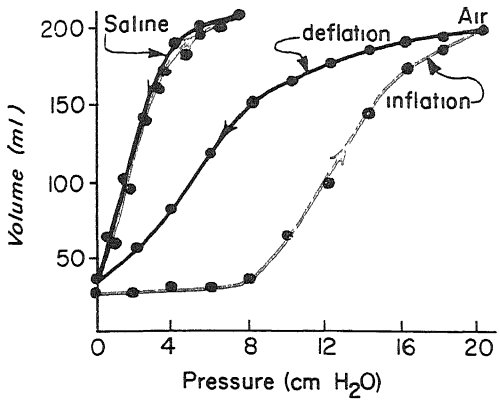


Fig 6 20 Air versus saline filling excised cat lungs at room temperature Saline curves obtained from excised gas free cat lungs suspended in saline bath at room temperature and filled from a saline reservoir Air curves taken from the gas-free state obtained with the lungs hanging in air and filled from an air reservoir maintained at a constant pressure Volumes all taken relative to the initial volume of the lungs as removed from the animal During inflation with air the volume equilibrated slowly and the lungs filled irregularly During deflation equilibrium was reached rapidly and the lungs emptied uniformly (From Radford 1957)

energy which is unaccounted for (hysteresis), i e , at any volume the product of volume and pressure is less on expiration than during inspiration The explanation for this "loss" is unclear It is related, however, to a force, surface tension, present at the interface of air and the fluid lining the alveolar walls This is clearly shown in Figure 6 20 in which the normal hysteresis of air-filled lungs is seen to be virtually eliminated by substituting saline for air Some of the concepts relating to this important observation are the following

The lung is comprised of several million more or less polygonal or spherical alveoli interconnected by tubes It is tempting, therefore, to assume that some properties of the lung compare to those of a physical model made of a T-tube with an inflated bubble or balloon at each end of the top of the T If the bubbles or balloons are inflated so that the pressure (P) is uniform throughout and if the stem of the T is then clamped off, one balloon or bubble may empty into the other rather than remain inflated The reason for this instability of the system is explained by the relationship between pressure (P), intramural tension (T), and radius of curvature (r) of curved surfaces This is stated in the LaPlace equation $P = 2T/r$ If one assumes that at inflation the tension within the walls of the bubbles and the radii of curvature are precisely equal, the bubbles will both remain inflated If, however, the radius of one bubble decreases slightly, the pressure in this one will rise if intramural tension remains constant and air will flow into the other bubble whose radius will increase, lowering its pressure (Fig 6 21) This phenomenon is not normally seen in the lung, i e , once the lung is inflated, it remains so throughout rather than collapsing in one area (atelectasis) and over-inflating in another

The stability of the lung is partly due to the fact that intramural tension (T) is small for the surfaces of the alveoli This property is attributed to a substance in the fluid lining the air spaces which reduces surface tension (Pattle, 1963) The bubble of fluid which opens up into the airway is, therefore, low in its intramural tension and therefore low in its tendency to contract or deflate

A substance which reduces the surface tension is a normal constituent of lung tissue extracts (Clements, 1957) Such extracts have a surface tension approximately $\frac{1}{10}$ the surface tension of water and $\frac{1}{40}$ the surface tension of other body fluids Phospholipids or lipoproteins have been implicated, lecithin being specifically identified among this group of compounds

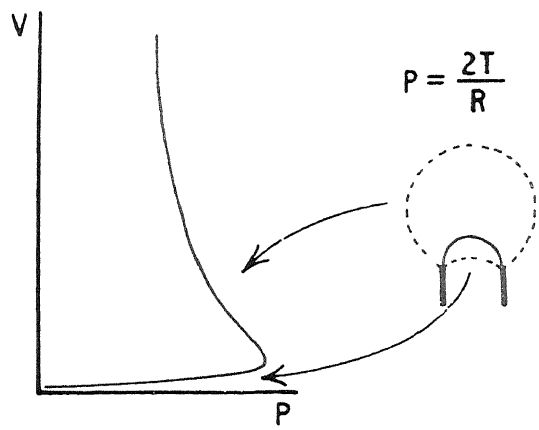


Fig 6 21 Volume pressure curve for a bubble on a tube As pressure rises volume changes little until a critical point is reached at which volume increases without further pressure rise This point is reached when radius of curvature (R) passes through a minimum value equal to the radius of the tube Thereafter at a constant tension (T) in the wall of the bubble pressure falls as radius or volume of the bubble increases Alveoli of the deflated lung pop open when the lung is gradually inflated as though following this law (From Mead 1962)

Extracts from lungs of children suffering from inability to maintain the inflation of their lungs (respiratory distress syndrome) are deficient in this surface active agent (Avery and Mead, 1959)

Not only do the surface active materials decrease surface tension, but they decrease it more when surface area is small than when it is large Surface tension is consequently less in small alveoli than in large ones Assuming alveoli change in size during inspiration and expiration, surface tension is less at the end of expiration than at the end of inspiration This decrease in surface tension with deflation and size balances the increase in pressure one would anticipate from the LaPlace equation if surface tension remained constant and radius decreased Stability is promoted, therefore, both because surface tension is generally reduced and because it adjusts itself to surface area

It appears, therefore, that the mechanical properties of alveoli reflect those of at least two concentric layers one is a film characterized by surface activity which normally stabilizes the air spaces of inflated lung and has its own pressure volume characteristics on inflation and deflation, the other is the alveolar epithelium together with the collagen, elastin, and other tissue possessing its independent mechanical attributes

THE FIRST BREATH

Breathing motions of a sort go on *in utero*, for material injected into the amniotic fluid can be recovered from the lungs after birth Therefore, the lungs contain fluid before the first breath The following are some of the mechanical considerations pertaining to their dramatic change to organs of respiration

To the extent that the spaces of the lung are distended with fluid, the pressure required for expansion conceivably is less than if they were collapsed, for some of the smaller airways will have a larger radius and, by the LaPlace equation, will enlarge with less pressure In point of experimental fact, lungs containing some fluid expand at a lower pressure than airless lungs containing no added fluid

The chest is compressed by pressures up to 95 cm of H_2O during vaginal delivery Some fluid is expressed from the lungs by this force in babies born with cephalic presentations The balance of fluid must pass into the circulation either directly across the alveolar epithelium into the pulmonary capillaries or indirectly through the lung lymphatics into the systemic circulation

The change from the airless to air-filled state is precipitous, it occurs in less than 1 sec after onset of breathing (Fawcitt et al , 1960) The functional residual capacity at 10 min of age is about 17 ml /kg , is about 30 ml /kg after 30 min , and is at this same value 4 days later Interpleural pressures recorded through an esophageal balloon during the first inspiration vary from -10 to -70 cm of H_2O The first expiration is associated with positive interpleural pressures as high as 20 cm of H_2O (Karlberg et al , 1962) Presumably these pressures are the result primarily of diaphragmatic and abdominal muscle contraction

4 / CONTROL OF BREATHING

CONTROL OF INSPIRATION AND EXPIRATION

In this chapter first the neural mechanisms which are responsible for rhythmic inspiration and expiration are described This is followed by analysis of those mechanisms which influence the rate and depth of breathing in such a way as to keep the carbon dioxide tension and oxygen tension of the arterial blood within the ranges compatible with normal functioning of the body At present most of the essential facts about the neuromuscular machinery for breathing and the sensors and stimuli which are concerned with regulating alveolar ventilation are known

As described in chapter 3, inspiration during eupnea is accomplished by contraction of certain muscles, such as the diaphragm and interosseus external intercostals, which may be called *primary* muscles of inspiration, while expiration is the result of the elastic recoil of the lungs and thoracic cage, and the action of gravity when the inspiratory muscles relax With increased demands for ventilation, additional muscles, collectively called *accessory* muscles of inspiration, are activated to produce a more rapid and deeper inspiration, certain muscles also contract during expiration to accelerate the return of the chest to the resting end-expiratory position In eupneic breathing, therefore, for inspiration and expiration to occur it is necessary simply for the motor neurons supplying the primary muscles of inspiration to show intermittent bursts of activity

The initial observation which is important in considering regulation of breathing is that apnea occurs when the CO_2 tension of the blood is below certain levels if at the same time the O_2 tension is above certain levels This situation can be produced simply by hyperventilating with atmospheric air, or an even longer period of apnea can be produced by hyperventilating with oxygen After hyperventilation is terminated apnea will supervene until CO_2 tension becomes elevated to a certain level or until O_2 tension becomes lowered below certain levels Thus, it is seen that the initiation of breathing after hyperventilation depends upon stimuli which derive either from a rise in CO_2 tension or from a fall in O_2 tension or from an interaction of these changes

The relative importance of CO_2 and O_2 in influencing breathing in an individual at sea level (ambient pressure of 760 mm Hg) is illustrated when the percentages of these gases in inspired air are altered If the O_2 percentage is increased above the amount normally found, even up to 100%, negligible influences on rate and depth of breathing are observed Also, if the amount of O_2 in inspired air is moderately lowered, namely, down to about 18%, typically, little or no change in breathing is observed Some persons will show stimulation of breathing as the O_2 is lowered to around 16% and levels of 10 to 12% regularly cause hyperpnea From these facts it is apparent that reduction in O_2 tension is not an important factor in regulation of breathing until the tension is reduced below some critical level which is somewhat lower than that found in normal resting subjects breathing air at a pressure of one atmosphere Any slight increase in the CO_2 tension of arterial blood, on the other hand, causes an increase in breathing, and a decrease in CO_2 tension is associated with apnea even when the O_2 tension is moderately below that which is found normally at sea level Such facts provide a basis for the interpretation that the inspiratory process at sea level in normal subjects is related to the action of CO_2 on certain neurons in the respiratory center in the medulla oblongata, causing them to discharge, and the termination of inspiration, resulting in expiration, occurs as a result of "feedback" of inhibitory influences to these neurons This mechanism serves to keep the CO_2 tension of arterial blood at a constant level of about 38 to 40 mm Hg in the presence of wide fluctuations in O_2 tension of arterial blood as long as the latter are at levels from about 80 to 90 mm. Hg and up

It is customary and convenient to speak of the collections of nerve cells in the brainstem which are concerned with the integration of the activity of the muscles of respiration as the *respiratory center* But the use of the term should not be taken to imply that the controlling nervous elements are a compact circumscribed mass, or confined to a closely restricted area Section through the brain at any level rostral to the upper border of the pons usually does not alter significantly the respiratory rhythm But sections at various levels behind this cause pronounced disturbances in respiration If a cut is made through the medulla caudal to the tip of the calamus scriptorius, all breathing ceases

Legallois (1824) located the respiratory center in the lower part of the medulla oblongata, and Flourens (1851, 1858) found a small spot about the size of a pin's head just beneath the forepart of the *calamus scriptorius* in the floor of the fourth ventricle, on either side of the midline He showed that after bilateral destruction of this area, which he named *noeud vital* (vital knot), the breathing ceased It appears, however, that the respiratory failure in these animals could have been secondary to circulatory failure related to loss of vasomotor tone

Markwald (1887) described powerful and prolonged tonic inspiratory movements or "cramps" which supervened after bilateral section of the vagus nerves and division of the brainstem immediately behind the posterior colliculi He concluded that a center inhibitory to inspiration was located in the latter situation, but that the vagi also had an inhibitory action, consequently the inspiratory "cramps" appeared only after vagal influence had been abolished as well Markwald's observations were confirmed shortly afterward by other experiments The subject was studied further by Lumsden (1923), who found that the prolonged inspiratory movements occurred only if the section passed through the pons a few millimeters behind its anterior border and occurred whether or not the vagi were divided The inspiratory cramps or *apneuses*, as Lumsden preferred to call them, last for several seconds He postulated their dependence upon an *apneustic* or *inspiratory center* at the level of the striae acusticae, which was dominated normally by an inhibitory or *pneumotaxic center* situated in the upper part of the pons The duty of the latter center was, through its inhibitory influence, to help to transmute the apneustic movements into the rhythmical movements characteristic of normal respiration After section of the brainstem behind the striae acusticae, the respirations consisted of a series of gasps occurring at relatively long intervals Lumsden concluded that these represented the activity of a primitive *gasping center* situated in the lower part of the medulla from which the two higher centers had evolved It was considered to be of little importance in higher animals

Lumsden's results were confirmed in the main by Stella (1938) and by Pitts, Magoun, and Ranson (1939) Stella, however, found in contradiction to Lumsden, that section through the pons (i e , separation of the pneumotaxic center) did not result in apneusis unless the vagal influence was abolished also (see Fig 6 22) The results of Pitts and his associates are in essential agreement with those of Stella They found that animals decerebrated through the upper part of the pons maintained a normal type of respiration which responded in the usual way to chemical and peripheral nerve stimulation so long as the vagi were intact, but immediately developed apneustic respiration and a complete cessation of rhythmical movements when both vagi were severed Stimulation of the central end of one of the cut vagi temporarily restores the respiratory rhythm The apneustic center is, therefore, under a double inhibitory influence, either one of which is capable of converting the apneustic type of respiration to the rhythm of normal or nearly

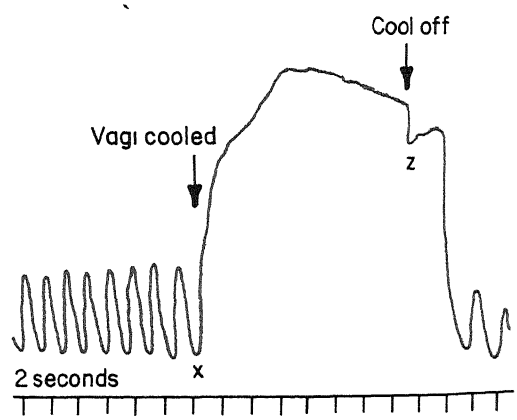


Fig 6 22 Showing apneusis Section of the brainstem along a plane passing dorsally immediately behind the posterior colliculi and ventrally 2 5 mm below the upper border of the pons Between *x* and *z* the vagi were blocked by cold Time 2 sec (After Stella 1938a)

normal respiration The vagal impulses influencing the apneustic center are initiated by the stretch of the lung toward the latter part of the inspiratory phase of normal breathing Studies by Cohen (1958) indicate that the pneumotaxic center shows rhythmic activity synchronous with breathing when its connections with the respiratory center are intact but fails to show rhythmicity when these pathways are cut

The apneuses, like normal respirations, are affected powerfully by the CO_2 tension of the blood, being increased in depth by having the animal breathe an air mixture containing a high concentration of CO_2, and reduced in depth, or prevented from occurring, by carbon dioxide lack According to Stella, the pneumotaxic center is bilateral but its connections with the apneustic center are mainly homolateral, i e , uncrossed

Pitts, Magoun, and Ranson (1938) describe the respiratory center in the cat, which they locate in the reticular formation of the medulla, as consisting of an inspiratory and an expiratory division, both centers are bilaterally represented The *inspiratory center* occupies the rostral half or two-thirds of the reticular formation overlying the olivary nuclei on both sides, i e , beneath the caudal third of the floor of the fourth ventricle (Fig 6 23) When stimulated a maximal coordinated inspiration results, involving both diaphragm and thorax If stimulated during apneusis, the magnitude of the inspiratory movement is increased, if stimulated during an interval between apneuses, an apneustic movement is produced The expiratory center lies in the reticular formation dorsal to the inspiratory center Electrical stimulation within this area causes expiration, if stimulated during inspiration or during apneusis, these movements are inhibited Regular respirations—inspiration alternating with expiration—are induced by rhythmical stimulation of the inspiratory center, expiration then occurs passively Rhythmical stimulation of the expiratory center also produces regular respiration, spontaneous inspirations then alternating with the expiratory movements Intimate synaptic connections exist between the diffusely distributed neurons within each center and between the two oppositely acting centers on the same side, as well as between similarly acting centers on opposite sides of the medulla Thus, unilateral stimulation of a small area of the inspiratory center causes maximal contraction of all the inspiratory muscles Excitation of one center causes simultaneous activation of its fellow of the contralateral side, accompanied by reciprocal inhibition of both oppositely acting centers Thus, the alternating rhythm of inspiration is established and sustained, and the respiratory movements synchronized on the two sides of the body The functional importance of the bilateral connections is demonstrated in a striking manner by making a deep longitudinal cut through the midline of the caudal part of the medulla, then the respiratory rhythms on the two sides of the body become asynchronous (see Fig 6 24)

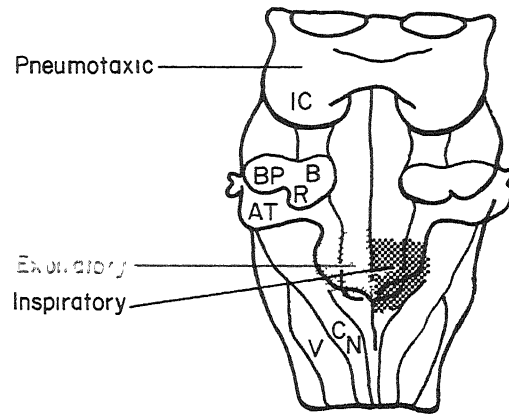


Fig 6 23 Dorsal view of lower brainstem of cat showing location of pneumotaxic inspiratory and expiratory centers Extent of expiratory center shown in light stippling inspiratory center in heavy stippling Though the centers are bilateral each is outlined on one side only for the sake of simplicity IC inferior colliculus AT acoustic tubercle B brachium conjunctivum B P branchium pontis R restiform body CN cuneate nucleus (Redrawn from Pitts Magoun and Ranson 1939a)

The similarity in the effects of stimulation of the expiratory center and of the central end of the vagus, namely, inhibition of inspiration and of apneusis, has led Pitts and his colleagues to the conclusion that the vagal respiratory effects are mediated through the expiratory center The influence of the pneumotaxic center probably is exerted in the same way

The respiratory center is connected with the motor neurons of the phrenic and intercostal nerves in the cervical (C 3, 4, and 5) and upper thoracic segments of the cord (T 2-6) by descending tracts which run in the anterior columns and in the ventral parts of the lateral columns of the spinal cord

For full descriptions of the neural mecha-

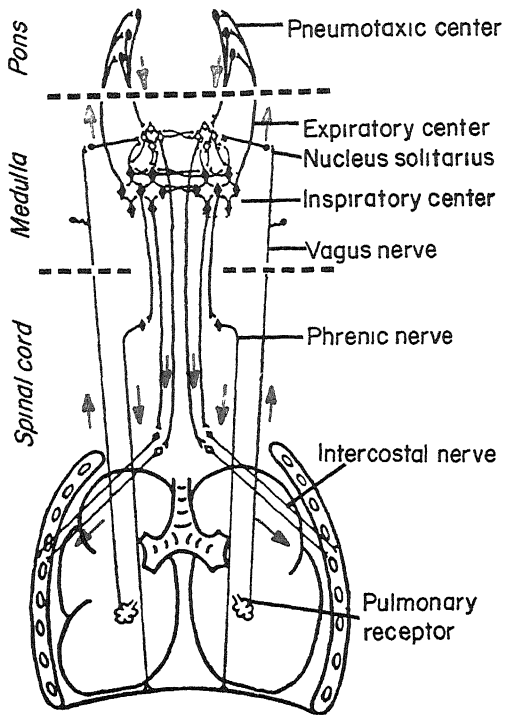


Fig 6 24 Diagram illustrating the chief nervous connections responsible for the control of the respiratory rhythm Inspiratory and expiratory centers in the medulla on the right side surrounded by dotted circles (After Pitts 1946 slightly modified)

nisms concerned with controlling inspiration and expiration the reader may refer to Wyss (1964) and Monnier (1968)

SPONTANEOUS ACTIVITY OF THE MEDULLARY RESPIRATORY CENTERS

This has been a subject of interest to physiologists for many years Some investigators, such as Coombs and Pike (1917), and Schafer (1918) have denied that the center is capable of spontaneous activity, maintaining that afferent impulses, especially those set up in the lung by the stimulus of stretch and conveyed by the vagus, and those initiated from proprioceptors in the respiratory muscles and traveling by the posterior nerve roots, were essential However, the results secured within recent years by means of improved methods of investigation indicate that, in certain species at least, the brainstem continues to discharge impulses to the respiratory muscles after all or nearly all afferent paths have been severed For example, rhythmic bursts of impulses can be recorded from the central stump of the phrenic nerve of a decerebrate animal after section of the vagi, glossopharyngeal and other cranial nerves entering the pons and medulla, and division of the spinal cord below the level of the seventh cervical segment Such an extensive operation would certainly interrupt all important afferent paths including those from the carotid sinus and the aortic arch It is not to be supposed, of course, that the respirations would be normal after such a radical procedure, for even if not essential for maintaining the activity of the center, afferent nerve influences are of the utmost importance in the regulation of that activity and the production of the normal respiratory rhythm Evidence for automaticity of the respiratory center of a cold-blooded species has been secured by Adrian and Buytenijk (1931) They succeeded in recording rhythmical action potentials from the excised brainstem of the goldfish, the potential changes had the same range of frequency as the respiratory movements (Fig 6 25)

The spontaneous respiratory activity is apparently dependent primarily upon the inspiratory center, the expiratory center playing an inhibitory role to interrupt intermittently the inspiratory discharge According to Cohen (1958) impulses are not discharged spontaneously from the pneumotaxic center Pitts and his colleagues suggest that the pneumotaxic center is excited from the inspiratory center, that a proportion of the impulses discharged from the latter region ascend to the pontine center whence impulses descend to the expiratory center, a discharge of *inhibitory* impulses is transmitted from the latter to the inspiratory center As the discharge of impulses from the expiratory center ceases, the inspiratory center resumes its activity, and the cycle is repeated A rise in the rate of discharge of the inhibitory impulses will increase the rate of breathing, while a reduction in the frequency of the dis-

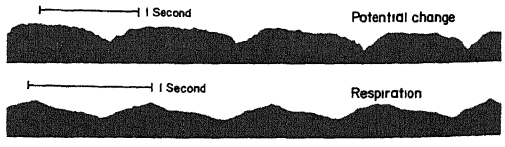


Fig 6 25 Comparison of electrical potential wave rhythm and rhythm of respiratory movements (After Adrian and Buytendijk 1931)

charge will be followed by slowing of the respirations

It appears, then, that in the intact animal the automatic alternating rhythm of the medullary centers is maintained largely by inhibitory impulses from the pneumotaxic center and from vagal afferents impinging upon the expiratory centers, which in turn inhibit, intermittently, the activity of the inspiratory centers The vagal influence is evoked by inflation of the lungs (i e , by inspiration), expiration, a passive act, follows (chap 3) Under ordinary physiological conditions the vagal regulating influence appears to play the dominant role

Of the two main methods which have been used for localizing inspiratory and expiratory centers, namely stimulation vs "sounding" for action potentials, the latter would seem to be somewhat more suitable Stimulation serves to expose only a dominant or major influence Neurons having opposite effects could be present and could be stimulated simultaneously, but the effects only of the dominant one of the two groups would be expected to appear Hence, inspiration elicited from a given site would not rule out the possibility that expiratory neurons also are present in this region, nor could an expiratory response elicited from a specific site prove that no inspiratory neurons were being stimulated simultaneously According to Brookhart (1940) lower strengths of stimuli applied to the regions in which the inspiratory center and expiratory center are supposed to be located will elicit various types of effects, and somewhat stronger stimuli are required to produce the effects reported by Pitts and co-workers On the other hand the recording of a burst of impulses only during one phase of the respiratory cycle would constitute evidence that only the inspiratory or expiratory neurons, depending on the time relations, were activated at that site under the conditions of the experiment When the "sounding" method is used, according to Gesell, Bricker, and Magee (1936), for the most part, impulses are recorded from a given site during both inspiration and expiration This would indicate that expiratory and inspiratory neurons are intermingled and would tend to make one doubt that there are anatomically discrete expiratory and inspiratory centers

A difficulty with the "sounding" method for determining the location of inspiratory and expiratory neurons of the respiratory center is that impulses may be picked up directly from motor neurons Achard and Bucher (1954) investigated the region of the respiratory center with needle electrodes and determined by subsequent histological examination the precise

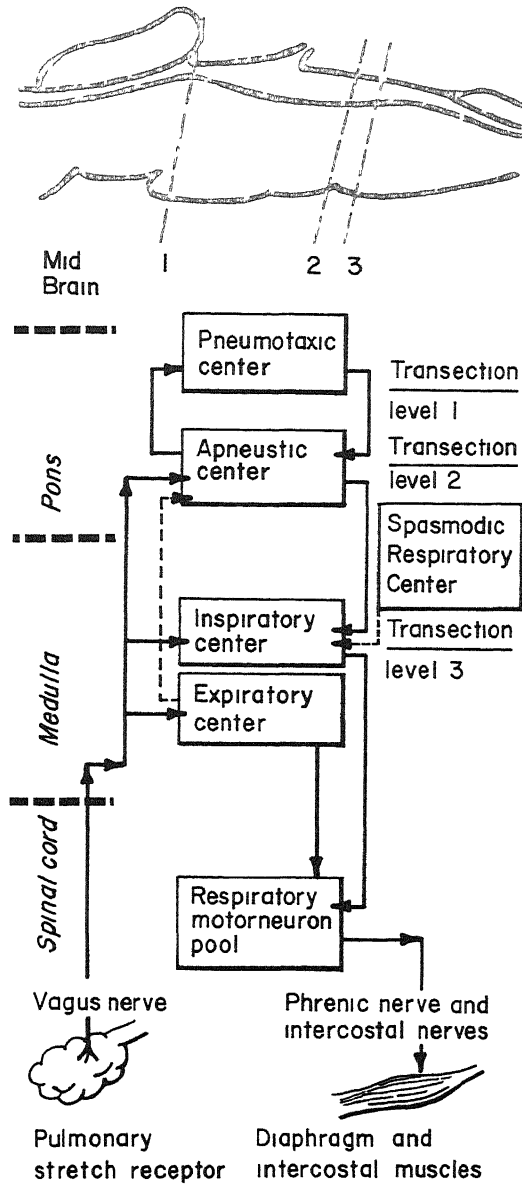


Fig 6 26 Schematic representation of organization of central respiratory mechanisms in the brainstem of the cat Key to transection levels is indicated above In vagotomized cats transection rostral to level 1 allows eupnea to be present transection between levels 1 and 2 causes apneustic breathing to appear transection between 2 and 3 results in eupneic breathing or gasping or occasionally Biot s breathing and transection caudal to level 3 is followed by absence of respiration (Reproduced by permission from Wang Ngai and Frumin 1957)

cells from which recordings were obtained Any purely inspiratory activity obtained by them was from *motor neurons* involved in inspiration, while action currents recorded from numerous other electrodes located in the area of the respiratory center did not permit a distinction between respiratory rhythms and the continuous asynchronous "background" activity in the reticular substance It must be concluded, therefore, that the question of the precise location of inspiratory and expiratory neurons within the medulla has not yet been answered

The relative importance of the pneumotaxic center in maintaining rhythmic breathing after vagotomy seems to differ according to the species of animal studied and the conditions of the experiments According to Hoff and Breckenridge (1954) in dogs, and under certain conditions in cats, midpontine section combined with bilateral vagotomy does not cause cessation of breathing, this has been confirmed by Wang and co-workers (1957) In fact vagotomy has been reported to have no effect on breathing in some instances in animals having the brainstem sectioned in the midpontine region Hoff and Breckenridge have returned to the view that the cause of the rhythmic discharge of respiration lies within the medulla itself "Just how the rhythm of respiration develops within the medulla is still problematical, and it may well be possible that within the confines of the medulla itself anatomical pathways are to be found wherein reciprocating self-limiting circuits, akin to those postulated between medulla and pons in earlier hypotheses, determine the rhythm of breathing "

The situation at present can be summarized by stating that the presence of a respiration-regulating group of neurons in the upper pontine region, the so-called pneumotaxic center, has been clearly demonstrated and has an important role in regulation of rhythmic breathing in the intact animal However, it seems to be clearly demonstrated, also, that rhythmic breathing still can occur when both this mechanism and the vagal reflexes are eliminated Finally, the precuse manner in which the pontine elements influence respiration has not been determined In Figures 6 26 and 6 27 are shown two recent attempts to explain the organization of the respiratory center

Profound inhibition of breathing can be produced by stimulation of the anterior cingulate cortex on the medial surface, and the orbitoinsulotemporal polar cortex on the ventral surface

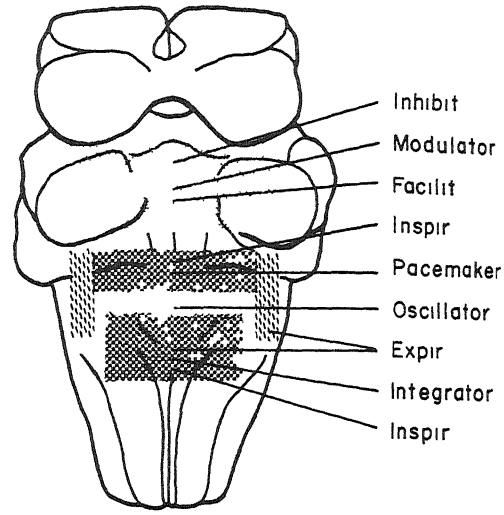


Fig 6 27 Schematic representation of functional concept of central regulation of respiration Relative locations in the brainstem of a cat of the regions which subserve the various functions are indicated (Reproduced by permission from Brodie and Borison 1957)

of the cerebral hemisphere In man, arrest of breathing has been produced in patients under light Pentothal anesthesia from the anterior cingulate and posterior orbital surface, and in the conscious patient from the same areas and from the anterior insula and the ventromedial aspect of the temporal pole Acceleration of respiration has been observed in carnivores on stimulating the motor cortex, the anterior ectosylvian and the sylvian gyri, and the middle and anterior portion of the cingulate cortex Precisely what role these cortical areas play in the control of breathing is not clear at present (See *Handbook of Physiology*, sect 1, vol 2 p 1347-1351)

NATURE AND LOCATION OF CO_2-SENSITIVE CELLS IN THE MEDULLA

That cells in the respiratory center show increased rates of discharge in response to an increase in CO_2 tension or hydrogen ion concentration in them, or in their immediate vicinity, is generally accepted The precise mechanisms of central sensitivity to changes in cH and Pco_2 however are not known One may recall that before the discovery of carotid bodies (which anatomically are not very far from the respiratory center) it was believed that oxygen lack had a stimulant effect directly in the respiratory center, and it is known now that such effects are on a reflex basis

Leusen (1954) showed that an increase in the hydrogen ion concentration in cerebrospinal fluid perfusing the cerebral ventricles stimulated ventilation, and he believed that this effect was exerted directly upon the respiratory center However, Loeschcke (see Grodins, 1964) showed that local application of acid to the floor of the fourth ventricle failed to stimulate breathing, hence he postulated the existence of a cerebrospinal fluid (CSF) bathed chemosensitive area outside the respiratory center Subsequently, Mitchell et al (1963) reported that specific areas on the ventrolateral surface of the medulla (bounded medially by the pyramidal tracts, laterally by the roots of the 8, 9, 10, and 11 cranial nerves, rostrally by the pons, and extending caudally about 6 mm) show chemosensitivity similar to that exhibited by the receptors of the carotid bodies Application to this area of pledgets containing CSF which was equilibrated with high P_{CO_2}, or having a high cH, or containing nicotine or acetylcholine, produced hyperpnea, whereas local cooling or application of procaine caused inhibition of respiration The authors postulate that all medullary CO_2 sensitivity is mediated via these "receptors" on the ventrolateral surface of the medulla The idea of specific chemoreceptors situated on the ventrolateral surface of the brainstem and influenced directly by changes in the cerebrospinal fluid has been further supported by Loeschcke (1965), Florez and Borison (1967), and Schlafke and Loeschcke (1968) Leusen (1969) states that agreement among researchers on the localization of central chemosensitive structures has not been reached, and Pappenheimer et al (1965) and Fencl et al (1966) have presented arguments against the view that there are such receptors located on the surface of the brainstem Leusen (1969) suggests a compromise interpretation, namely that both superficial and more deeply located zones are normally active but that the deeper zones are more sensitive to anesthesia and oxygen lack so that in anesthetized or shocked animals the relative importance of the superficial medullary chemoreceptors in maintaining breathing would be increased If this is correct the role of the superficially located chemoreceptors would be somewhat similar to that of the peripheral chemoreceptors of the carotid and aortic bodies (p 6-55) Evidence which supports this view is presented by Cozine and Ngai (1967), who failed to obtain effects on respiration by applying procaine to the ventrolateral medullary chemoreceptors except when the deeper structures were depressed by anesthesia or anoxemia

Reflex and Motor Mechanisms

HERING-BREUER REFLEXES

The Hering-Breuer reflexes are effects on the respiratory cycle which are elicited by inflation and deflation of the lungs The receptors concerned are in the lungs, and the afferent pathways are in the vagus nerves The function of these reflexes apparently is to regulate the respiratory cycle rather than to alter significantly the pulmonary ventilation for any prolonged period The observations made by Hering and Breuer (1868) were that inflation of the lungs causes arrest of inspiration expiration then ensuing, and deflation of the lungs has an inspiration-exciting effect They found that these effects were abolished when the vagus nerves were sectioned Another way to demonstrate the effects of lung distention is to block the airway at the beginning of either inspiration or expiration When inflow of air is prevented, a stronger inspiratory effort results than in the preceding cycle, and when egress of air is blocked at the end of inspiration, a longer interval than the preceding one occurs before the onset of the next inspiratory effort In eupnea apparently the inflation effect during inspiration alone is a factor As inspiration proceeds, distention of the lungs occurs, and impulses are carried by vagal afferents to the respiratory center where they cause inhibition of the inspiratory neurons, thus terminating inspiration .Once the lungs begin to deflate, impulses are no longer set up in the vagal afferents and, consequently, the inspiratory neurons again are free to discharge

The mechanical problem in studying the Hering-Breuer reflexes can be circumvented in the case of the rabbit since this animal has a slip of muscle on the posterior surface of the diaphragm which can be dissected free and recorded from, separately from the remainder of the diaphragm This diaphragmatic slip of muscle can be used as an indicator of the degree of activation of inspiratory muscles If the lungs are distended, the rate and amplitude of contraction of the slip decrease, and if the lungs are deflated, an increase in rate and amplitude of contraction of the slip is observed (Fig 6 28)

The stimulating effect of lung deflation on

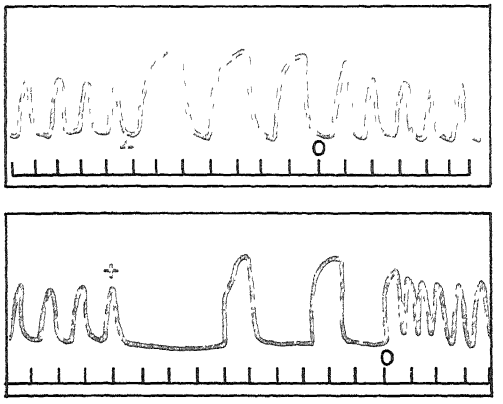


Fig 6 28 Record of contractions of isolated slip of a diaphragm of a rabbit to illustrate Hering-Breuer reflexes Upstroke is produced by contraction of the slip i e inspiration Above at + the trachea occluded at the end of expiration and at 0 the obstruction was removed Below at + the trachea was occluded at the end of inspiration and at 0 the obstruction was removed (After Head)

inspiration is related under ordinary conditions to failure in the production of the normal amount of inspiration-inhibiting influence at the end of inspiration However, severe deflation, as from pneumothorax, results in activation of other receptors, and reflex stimulation of inspiration follows Finally, if lung inflation is severe, not only is inspiration inhibited but active expiration is elicited reflexly This may not require a third type of receptor, but may be simply the result of greater degree of activation of the same kind of receptors which are concerned in the reflex inhibition of inspiration The role of vagal reflexes in regulating inspiration and expiration has been described in detail by Wyss (1954) Paintal (1963) reviewed what is known about the location and response patterns of pulmonary receptors

A physiological role of the Hering-Breuer reflexes is to regulate the extent of lung inflation so that the tidal volume tends to fall within a certain range Also, these reflexes serve to insure filling and emptying of the lungs despite a considerable increase in resistance in the airway As would be expected, the typical result following bilateral vagotomy is an increase in the depth of breathing The inspiratory act under these conditions is seen to be unnecessarily long since the chest tends to remain in the inspiratory position for a short period after ingress of air has stopped This is an inefficient process since muscular work is being done while no air is being moved. Also, a decrease in respiratory rate develops in most animals after vagotomy This probably is a compensatory change, secondary to the increased depth, which results in an amount of pulmonary ventilation similar to that which was present before vagotomy An increased depth of breathing without a decrease in rate would soon wash out CO_2 so that the stimulus to the respiratory center would be removed, hence the rate soon would decrease to cause restoration of normal CO_2 levels It should be mentioned, however, that the effects of vagotomy on rate and depth of breathing vary greatly depending upon the anesthetic agent which is used, and furthermore, changes are progressive for a prolonged period following vagotomy

The effects on respiration of direct stimulation of the central end of the vagus nerve sectioned at the cervical level are somewhat unpredictable, since the nerve contains several types of afferent fibers which carry impulses to the respiratory center These include *1)* aortic arch afferents, *2)* aortic body afferents, *3)* inspiration-inhibiting fibers activated normally during lung inflation, *4)* inspiration-exciting fibers activated physiologically by excessive lung deflation, and *5)* fibers concerned with visceral pain, and possibly still other types of fibers However, strong stimulation of the central end of the vagus nerve usually causes apnea, and this may persist for a few seconds after cessation of stimulation This result could be elicited if the effects of stimulation of *1* and *3* above were sufficient to override the effects of stimulation of *2* and *4*

The Hering-Breuer reflexes are basically proprioceptive in nature In this case contraction of skeletal muscle not only increases the tension in muscles and shortens them, but elastic objects, the lungs, also are put under greater tension The lungs also are supplied with receptors from which information is carried back to the same neurons that activated the muscles The muscles of respiration, like skeletal muscle in general, are supplied with proprioceptive innervation through which their contraction is regulated in the same way as described for the limb musculature

REFLEX EFFECTS FROM BARORECEPTORS

A sudden rise in pressure in the sino-aortic zones elicits reflexly a decrease in rate and depth of breathing If the pressure rise is considerable, apnea may be produced Typically, adaptation is rather rapid so that breathing is

restored even though the new pressure level is maintained Part of the recovery of respiration is related to adaptation in the receptors themselves, but also it occurs as a consequence of the rise in P_{CO_2} during apnea The opposite change, a sudden fall in pressure in the sino-aortic zones, elicits hyperpnea Thus, any sudden change in blood pressure per se tends to elicit respiratory responses During rapid bleeding or following injection of vasodilator drugs, such as nitroglycerin or bethanechol (a choline derivative), respiration is stimulated (Fig 6 29), and the effect is prevented in considerable part by sino-aortic denervation (Schopp, Gilfoil, and Youmans, 1957) The respiration-stimulating effect of nitroglycerin which still remains after denervation can be partly or completely prevented if a device is used to prevent the fall in arterial blood pressure

The effects on respiration elicited from the baroreceptors probably are of importance in influencing the venous return (Youmans, 1958) Vasodilation promotes pooling of blood, especially in the abdominal viscera, and increased respiratory activity promotes the movement of blood from the abdominal portion of the inferior vena cava on into the thoracic portion The concept of the respiratory pump, as a factor in aiding venous return, introduced by Donders in 1859, seems accurate according to the studies reviewed by Brecher (1956) Conversely, the respiratory inhibition during a rise in blood pressure, especially in a person in the orthostatic position, would favor pooling of blood in the splanchnic region so that venous return would be decreased, thus cardiac output would be lowered, and the rise in blood pressure would be counteracted In short, one may interpret the respiratory responses from baroreceptors, along with effects on heart rate and arteriolar tonus elicited at the same time, as being concerned with the stabilization of arterial blood pressure However, characteristics of these reflexes are such that they are useful only in counteracting sudden changes of brief duration, they are overridden easily by other influences

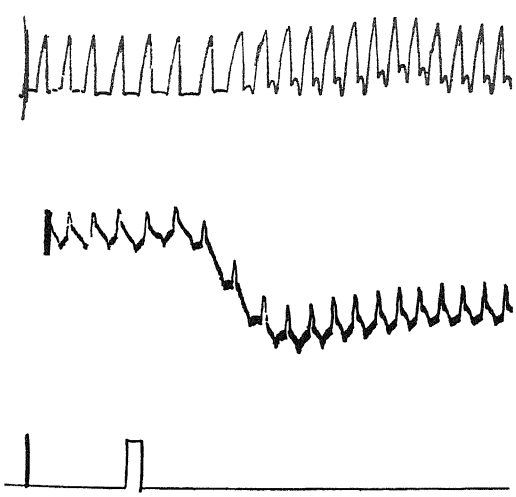


Fig 6 29 Effects on blood pressure (below) and respiration (above) following injection of 0 1 mg/kg of nitroglycerin into dog anesthetized with α-chloralose The respiratory stimulation can be prevented if the fall in arterial blood pressure is counteracted by means of rapid infusion of blood from a pressure-regulating device (From studies by Schopp Gilfoil and Youmans 1957b)

CONTROL OF ACTIVITY OF THE ABDOMINAL MUSCLES

In the resting, supine subject, abdominal muscles may be completely inactive throughout the breathing cycle (Campbell, 1958) In moderate hyperpnea they become active during expiration, and in more severe hyperpnea they show a rapidly augmenting activity during the early part of expiration, with a decrease well before the next inspiration begins Abdominal muscles usually are active during expiration in the subject in the standing position, and they contract in response to continuous positive pressure breathing in a subject in the supine position The degree of their response to positive pressure is proportionate to the level of pressure applied Gesell and Moyer (1941) reported that the contraction of abdominal muscles in response to positive pressure breathing in dogs did not occur after bilateral vagotomy at the cervical level Their discussion was based on the tacit assumption that vagotomy had this effect because of elimination of impulses from the lungs Apparently, no one had questioned the source of this reflex effect on abdominal muscles until recently Studies by the Wisconsin group (Youmans et al , 1963) revealed that positive intrapulmonary pressure elicits contraction of abdominal muscles in dogs with closed chest, but does not do so if the chest is open, even though lung distention in the latter case is greater, hence, Hering-Breuer reflexes apparently are not concerned Bishop (1964) likewise has concluded that the abdominal muscle response to positive pressure breathing "is not a simple corollary of the Hering-Breuer reflex but is a separate reflex served by its own vagal pathway "

Results of extensive studies (Youmans et al , 1963) indicate that activity of abdominal muscles may be increased selectively on a reflex basis when volume of blood in certain "low pressure" portions of the cardiovascular system is reduced Although it is clear that anything which stimulates breathing tends to cause appearance of, or an increase in, activity of abdominal muscles, on the other hand, a number of procedures, which lead to changes in amplitude of contractions of abdominal muscles, do so without causing corresponding changes in tidal volume In experiments on dogs it was found that several procedures (including bleeding, tilting the animal to the head-up position, aortic occlusion, and positive pressure breathing) cause marked changes in amplitude of steady-state contractions of abdominal muscles, while having little or no effect on tidal volume In Figure 6 30, a schematic representation is shown of main factors and mechanisms concerned with influencing the activity of abdominal muscles In addition to the well-known influences from chemoreceptors and baroreceptors, acting by way of the respiratory center, it is postulated that atrial mechanoreceptors (extensively studied by Paintal (1963) and others) are concerned with reflex inhibition of tonus of abdominal muscles Any decrease in volume of the atria is envisaged as causing an increase in tonus of abdominal muscles through withdrawal of this inhibitory effect The consequent rise in intra-abdominal pressure serves to reduce the capacity of the venous system and hence promotes better cardiac filling The influences from alterations in cardiac filling apparently can be exerted independently of the effects on amplitude of breathing (as shown for example by the fact that occlusion of the postcaval vein causes an immediate increase in activity of abdominal muscles, while occlusion of the ascending aorta causes immediate inhibition of such activity, yet minimal and similar effects on inspiration occur) Therefore, it is necessary to assume, as shown in Figure 6 30, that impulses from atrial mechanoreceptors can be channeled to the motor neurons which innervate the abdominal muscles without influencing the respiratory center The only new concept presented in the schema is that atrial mechanoreceptors (activated by increased atrial volume) are concerned with reflex inhibition of tonus of abdominal muscles Campbell (1958) suggested the possibility that changes in activity of abdominal muscles during the respiratory act may be determined in part by reflexes which do not operate through the respiratory center He further stated that some of the control of abdominal muscles may be for the purpose of modulating changes in abdominal pressure, because of the effects on circulation, and that receptors in the circulatory system, in either the abdominal or thoracic cavities, might be involved Bjurstedt (1953), more specifically postulated that tonus of abdominal muscles may be influenced reflexly from receptors in the circulatory system that are sensitive to changes in filling As to the question of an abdominal muscle tonus center (*ACC* in the schema), it is noteworthy that Gesell, Magee, and Bricker (1940) observed steady-state "expiratory" discharge in neurons in the medulla and suggested that these units may be concerned with posture rather than with exhalation More recently Haber et al (1957) recorded two types of "expiratory" patterns from neurons in the medulla, which they differentiate as early vs late The late "expiratory" discharges are similar to Gesell's steady-state type Since they occur at maximum frequency in the anesthetized dog well after the end-expiratory position has been reached (at a time when there is no movement of air and when abdominal muscular activity and intra-abdominal pressure are maximal) it would seem to be at least as appropriate to call them abdominal compression discharges as to call them expiratory discharges Haber et al (1957) also found that the neurons are carbon dioxide sensitive and may become inactive as a result of hyperventilation This is also a characteristic of the neuronal system which is concerned with maintaining steady-state activity in abdominal muscles (Youmans et al , 1963)

Among the best evidence for reflex influences on abdominal muscle tonus elicited from a cardiac chamber is that reported by Turner (1961), who demonstrated that distention of the left atrium produces inhibition of tonus of abdominal muscles in dogs being artificially respired and having the systemic and pulmonary circulation maintained constant by means of a pump This would seem to implicate the left atrium as at least one site from which tonus of abdominal muscles can be influenced, but it does not rule out the possibility of influences from still other sites in the heart or pulmonary circuit

REFLEXES INITIATED IN OTHER PARTS OF THE BODY

The experiments of Churchill and Cope (1929) showed that a reflex increase in the re-

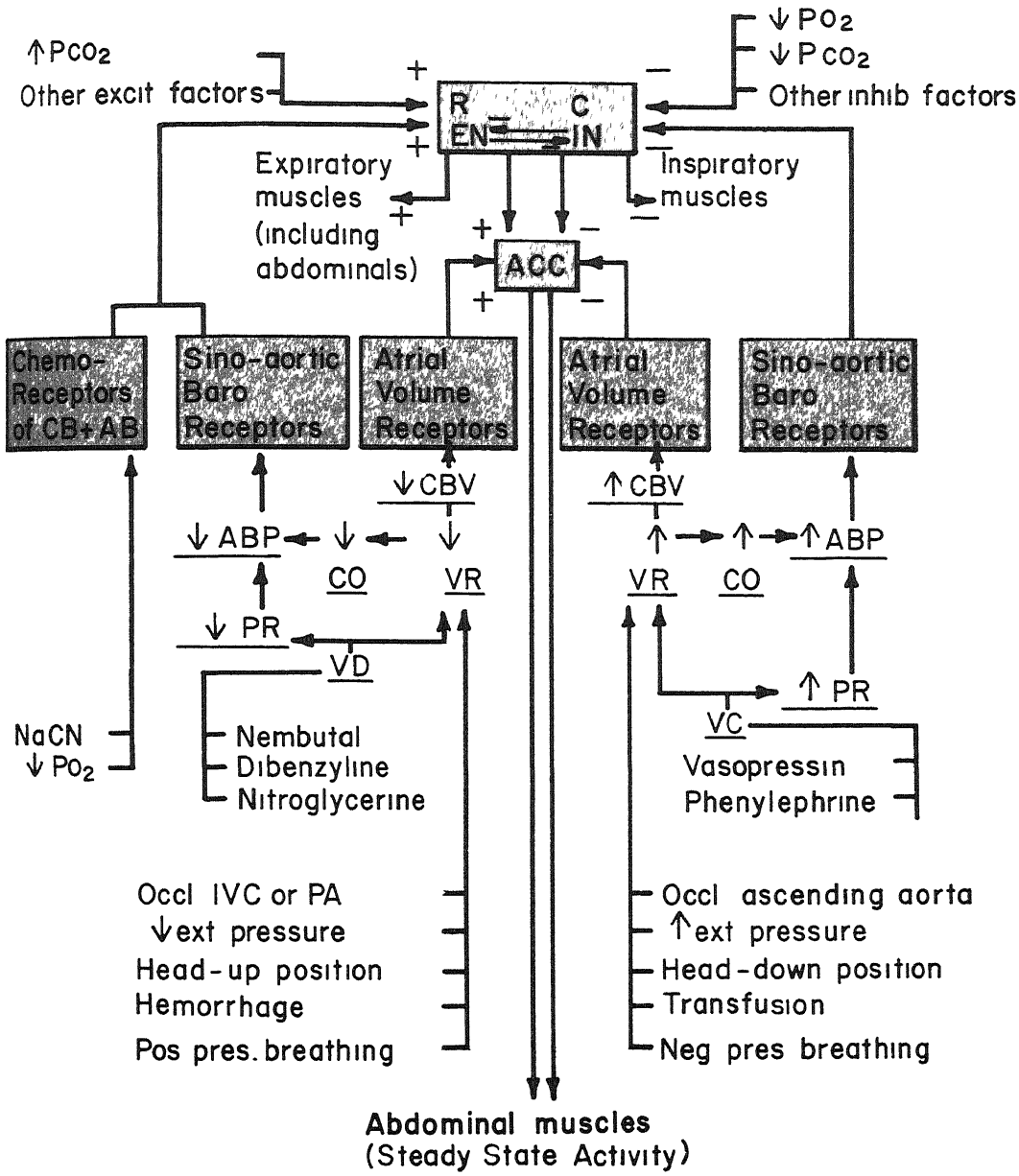


Fig 6 30 Schematic representation of main factors and mechanisms concerned with altering the level of activity of abdominal muscles in response to cardiorespiratory changes The left side of the diagram illustrates excitatory influences which affect the activity of abdominal muscles either through the respiratory center or more directly The right side shows opposite changes which promote a decrease in the activity of abdominal muscles *RC* the respiratory center *EN* expiratory center *IN* inspiratory center *ACC* abdominal compression center *CB* carotid body *AB* aortic body *CBV* central blood volume *ABP* arterial blood pressure *CO* cardiac output *VR* venous return, *PR* peripheral resistance *VC* vasoconstriction *VD* vasodilation *IVC* postcaval vein *PA* pulmonary artery + excitatory influence − inhibitory influence

spiratory rate results from distention of the pulmonary vascular bed When the vessels of the lung, which was isolated from the body except for its nerves, were injected with fluid to cause overfilling of the vascular bed, rapid shallow breathing immediately ensued Section of the vagal fibers abolished the response

Stimulation of almost any afferent nerve may bring about a reflex change in respiration Stimulation of pain fibers is especially potent

in this regard, and the respiratory effects of the excitation of the cutaneous nerves by extremes of heat or cold are well known The thermal effect upon respiration is apparent in the panting of the dog The increased pulmonary ventilation in fever is also partly due to the stimulation of receptors (*thermoreceptors*) responsive to a rise in temperature The thermoreceptors are situated peripherally, especially in the skin, and centrally in the hypothalamus, the latter, being stimulated through the rise in temperature of the blood, set up impulses which are transmitted to the respiratory centers, the pneumotaxic center in particular The great increase in pulmonary ventilation occurring in muscular exercise is dependent in part upon reflexes originating in the active muscles and moving joints Proprioceptive impulses from the diaphragm and other respiratory muscles during one respiratory movement exert an important influence upon the succeeding movement Stimulation of sensory nerves in the respiratory tract, as by ether anesthesia, or of the abdominal viscera, either during surgical operations or as the result of disease may cause profound changes in breathing Also, as pointed out by Harrison and his associates (1932) stimulation of afferent nerve endings in the great veins and right auricle by the rise in venous pressure is a possible factor in the hyperpnea of exercise or in the dyspnea of cardiac failure The glossopharyngeal nerve contains afferent fibers which inhibit respiration during the second stage of the act of swallowing Abrupt inhibition of respiration is also caused by the inhalation of an irritant gas through stimulation of nasal branches of the fifth nerve In other instances irritation of these endings may cause sneezing—a modified respiratory act Coughing is most commonly reflex in character, and may be initiated by the stimulation of afferent nerve endings in the trachea and larynx

Also, there are receptors in the heart, which notably are activated by veratridine injected into the coronary arteries, and elicit inhibition of respiration—along with a decrease in heart rate and inhibition of arteriolar tonus The physiological role of these receptors is not known, but stretching of the cardiac chambers by overfilling elicits a similar response and it is possible that these receptors are concerned in this reflex.

Regulation of Pulmonary Ventilation

RESPIRATORY ADJUSTMENTS TO CO_2

Pulmonary ventilation in a normal subject at rest is kept relatively constant at a given level in response to the carbon dioxide tension, hydrogen ion concentration, and oxygen tension in specific sites Thus, ventilation is such as to tend to keep the values for $P_{A_{O_2}}$, $P_{A_{CO_2}}$ and pH within a narrow range As a consequence of an increase in metabolic rate oxygen tension tends to be decreased and carbon dioxide tension and hydrogen ion concentration tend to be increased This does not happen to any considerable degree because changes in these chemical stimuli elicit an increase of pulmonary ventilation to counter a rise Or one can visualize that if the basal metabolic rate (BMR) suddenly should increase by a given increment the respiratory minute volume (RMV) would also be increased by feedback from sensors responding to the chemical stimuli so that the levels of $P_{A_{O_2}}$, $P_{A_{CO_2}}$, and pH would be kept virtually the same as before the rise In the case of responses to exercise, however, additional mechanisms are utilized to regulate RMV

Although it is generally accepted that the major chemical stimuli which are concerned with influencing pulmonary ventilation are known, there are a number of questions which require further study These include questions concerning sites and mechanisms of action, relative sensitivity in the different sites, and possible interactions between the stimuli

The CO_2 content of inspired air is negligible, as far as human physiology is concerned, and can be increased somewhat without causing an increase in RMV Its concentration at sea level can be increased by about 1% (7 6 mm Hg) before a measurable increase in pulmonary ventilation occurs If 4% CO_2 is breathed, the RMV is approximately doubled on the average, but the effect may be scarcely noticed by the subject Any further increase in CO_2 from 4% up to 10% produces a correspondingly greater RMV, and at the latter level it may be 8 to 10 times the resting level Also the subject may have symptoms of dizziness, faintness, and headache Increasing the CO_2 percentage beyond 10 does not result in significantly greater increases in RMV, and at levels above 20% the

RMV may begin to show a decrease Also, these levels cause convulsive seizures From the practical standpoint it would appear that if CO_2 is to be administered to stimulate breathing, there would be little use for any mixture containing less than 5% or more than 10% CO_2

It is generally agreed that the primary site of action of CO_2 in producing effects on breathing is in the respiratory center or in the immediate vicinity of the center Cells in a respiratory center which has been isolated from extrinsic neural influences still show alterations in rates of discharge in response to corresponding changes in CO_2 tension of the blood reaching them, and injection of bicarbonate directly into the respiratory center causes an increased activity of these neurons

Considerable attention has been given to the question of the precise mechanism by means of which CO_2 exerts its stimulant action It has been debated whether CO_2 itself has such an effect, or whether the effect is exerted indirectly as a consequence of the concomitant increase in hydrogen ion concentration There is evidence, analyzed by Gray (1950), that both a rise in P_{CO_2}, and in cH, in the respiratory center, contribute to and supplement each other in causing an increase in pulmonary ventilation.

Another mechanism, reflex in nature, is available for the stimulation of breathing as a consequence of a rise in the P_{CO_2} of arterial blood The chemoreceptors, already described as concerned with producing circulatory adjustments in response to changes in P_{O_2} and P_{CO_2}, when subjected to a sufficient rise in P_{CO_2} will elicit an increase in RMV The extensive literature dealing with chemoreceptors has been summarized by Heymans and Neil (1958) and others in the proceedings of the J S Haldane Centenary Symposium (Cunningham and Lloyd, 1963) The fact of reflex stimulation of breathing in response to rises in P_{CO_2} of the blood flowing to the aortic and carotid bodies has been demonstrated by a number of methods and is generally accepted A question which has been vigorously disputed, however, is whether the respiratory center in the intact subject is more sensitive than the chemoreceptors In other words, does the chemoreceptive reflex mechanism constitute the first line of defense against a rising P_{CO_2} in arterial blood or does the respiratory center have the greater sensitivity Heymans has supported the view that the chemoreceptors are concerned with reflex drive of breathing even under resting conditions at sea level, while Schmidt and Comroe have favored the interpretation that these receptors are activated in significant degree only after considerable increases in arterial P_{CO_2} The details of the argument are reviewed by Heymans and Neil (1958), who conclude that the question remains open as to how much the chemoreceptors participate in the response to moderate hypercapnia in the intact animal However, the majority viewpoint seems to be that the normally sensitive central mechanism responds to rising CO_2 and through increased ventilation prevents a rise in P_{CO_2} of arterial blood sufficient to elicit a response from peripheral chemoreceptors If, on the other hand, the respiratory center is rendered relatively insensitive to CO_2, for example, by anesthetic agents and narcotic drugs, respiration is depressed, and P_{CO_2} rises to a level sufficient to set up peripheral chemoreceptive drive of breathing According to the latter interpretation the CO_2-induced chemoreflex constitutes a secondary line of defense to help insure pulmonary ventilation when the sensitivity of the respiratory center is reduced

Although it would appear that influences from peripheral chemoreceptors are not always needed, and perhaps would not be a significant factor in eliciting an increase in RMV in response to a slowly rising CO_2 tension in a normal subject, these receptors are concerned in eliciting the initial response to a sudden rise in $P_{A_{CO_2}}$ It is reported that in an animal with peripheral chemoreceptors denervated and normoxic, so that no significant oxygen lack drive is a factor, the response to inhaled CO_2 is demonstrably different from that of the animal prior to denervation The denervated animal responds after a longer latent period and then more sluggishly (Leitner et al , 1965) This result is not surprising and does not necessitate the conclusion that impulses from peripheral chemoreceptors elicited by the normal CO_2 tension are concerned in driving breathing in the animal at rest Perhaps more applicable to the question is the claim that after chronic denervation of peripheral chemoreceptors the resting $\dot{V}_E$ of the unanesthetized animal is about 80% of what it was before denervation and the $P_{A_{CO}}$ is about 10 mm Hg higher (Mitchell et al , 1964, Bouverot et al., 1965) This appears to

indicate that about 20% of the drive during normal breathing at rest is derived from peripheral chemoreceptors

Ezaguirre and Lewin (1961) approached the question of sensitivity of the peripheral chemoreceptors directly by determining the number of impulses in afferent nerve fibers from the carotid body of the cat breathing different oxygen mixtures at various P_{CO} levels produced by different levels of ventilation If we may assume normal sensitivity of the chemoreceptors in their study it may be concluded from their results that impulses are discharged from peripheral chemoreceptors in response to a normal P_{ACO} (since a rise from 30 to 40 mm Hg produced a considerable increase in frequency) Also a reduction in oxygen content in inspired air from 50 to 20% resulted in a slight increase in the number of impulses set up in the nerve fibers from the carotid body The latter effect was slight enough, however, to indicate that the influence of oxygen lack on chemoreceptors at normal P_{A_O} is negligible

The lowering of P_{CO_2} in arterial blood can be achieved by voluntary hyperventilation or by excessive artificial ventilation of anesthetized subjects Following hyperventilation, apnea is observed, and breathing is resumed when either P_{CO_2} rises to approximately the level found in the subject under resting conditions or when the P_{O_2} of arterial blood drops to a level sufficient to set up reflex drive of breathing on an O_2 lack basis, whichever occurs first, After one has hyperventilated moderately with atmospheric air, oxygen lack develops quickly as the store of oxygen in the lungs is absorbed and utilized, breathing is resumed in response to the oxygen lack stimulus before the P_{CO_2} has been restored to the normal resting level Then the oxygen lack is relieved by a few breaths, and apnea appears again Gradually, the P_{CO_2} level becomes restored and takes over as the sole drive This is the basis for the appearance of a few cycles of periodic breathing after hyperventilation As would be expected if one hyperventilates with 100% O_2, rather than air, just as much CO_2 is blown off as before, with a given degree of respiratory activity, but the P_{O_2} in alveolar air is quite high at the end of the period of hyperventilation Consequently a longer time will be required for the O_2-lack stimulus to appear than for the P_{CO_2} level to be restored Therefore, following hyperventilation with 100% O_2 the elevation of P_{CO_2} may be the initial factor in the restoration of breathing

RESPIRATORY ADJUSTMENTS TO O_2

The P_{O_2} in inspired air can be reduced by lowering the atmospheric pressure in a decompression chamber in which the subject has been placed, or he may breathe prepared gas mixtures having various ratios of O_2 to N_2 Since O_2 tension is lowered in proportion to the decrease in barometric pressure with altitude, one may indicate the altitude at which the experimentally induced P_{O_2} would be encountered Thus a certain altitude, as far as P_{O_2} is concerned, is simulated by the experimental conditions For example, if one breathes a mixture of 10 5% O_2 and 89 5% N_2 at 760 mm Hg pressure, the P_{O_2} is about one-half that in atmospheric air at the same pressure This approximates the P_{O_2} found in atmospheric air at an altitude of 18,000 ft where the barometric pressure is about 380 mm Hg The details concerning respiratory responses to P_{O_2} lower than that of room air at sea level have been described by Dripps and Comroe (1947) and by Rahn and Otis (1949) The major observations may be summarized as follows Some subjects show a slight stimulation of breathing when the O_2 is lowered to 16 to 18% while others do not respond until a level in the range from 16% down to 12%, or even lower in occasional subjects, is reached At levels of 10% or below, breathing is stimulated but irregularities may develop As the O_2 tension in inspired air is lowered and respiratory minute volume increases, the CO_2 tension in alveolar air and arterial blood falls Thus at the same time that the oxygen-lack stimulus is operating to cause an increased RMV, the CO_2 and cH drives of breathing are being withdrawn It is apparent, therefore, that the stimulating effects of oxygen-lack per se on breathing could be recorded if the fall in CO_2 tension were prevented, as by adding the appropriate amount of CO_2 to the inspired air In the discussion of hyperventilation it was noted that a relatively slight reduction in CO_2 tension (3 to 4 mm Hg) results in apnea when there is no accompanying O_2 lack At altitude, therefore, where the CO_2 tension is lowered more than this (secondary to increased RMV), it must be assumed that respiratory drive is entirely on an oxygen-lack basis or that a readjustment in sensitivity to

CO_2 tension has occurred Acutely, the former seems to be true, while after prolonged exposure to low oxygen tension the latter change apparently occurs

Papers on respiration written before the discovery of the chemoreceptors of the carotid bodies refer to a stimulant action of oxygen lack on the respiratory center It is now believed that a lowered O_2 tension in the medulla does not result in stimulation of breathing, and decreases in O_2 tension of arterial blood, within ranges that the animal can tolerate for significant periods, do not cause stimulation of breathing if the carotid and aortic bodies have been denervated Thus, the only sensitive mechanism which responds to oxygen lack and elicits an increase in RMV is the chemoreceptor reflex Here, too, as in the case of the CO_2 effect, the precise level at which a lowered arterial O_2 tension elicits stimulation of breathing is disputed Comroe and Schmidt (1938) first reported that the O_2 tension in blood perfusing the carotid bodies was lowered to about 50 mm Hg before the receptors were activated, but others have reported the threshold to be above 100 mm Hg Po_2 Experiments showing the effects on RMV of blocking the chemoreceptors do not serve to establish the level of activity under normal resting conditions if anesthetic agents are used, since, as already explained, these substances usually cause depression of the respiratory center so that some part of the drive of breathing is transferred to the chemoreceptors If oxygen lack stimulates breathing only through activation of chemoreceptors, then depression of breathing should be produced as a result of breathing mixtures containing more oxygen than that found in atmospheric air Loeschcke (1953) reported an 8% reduction of RMV for about a minute when 32% O_2 in N_2 was substituted for room air in unanesthetized man, and this was followed by a rise until equilibrium was reached after about 4 min On returning to room air, RMV increased a little for a few minutes and then returned to the control level Dripps and Comroe (1947), however reported only 3 1% reduction in RMV caused by inhalation of 100% O_2, whereas Baker and Hitchcock (1947) found an increase in ventilation in response to 100% O_2 The results cited indicate that if there is any drive of breathing related to O_2-tension effects in normal subjects at sea level it is so small that it can be detected only by very careful measurements Furthermore, it is possible that even the slight increases in ventilation which have been reported are not related to the removal of chemoreceptor drive, but rather to a rise in Pco_2 tension in the respiratory center secondary to the vasoconstriction produced by high O_2 tension in the arterioles of the brain (See Kety, S S and Schmidt, C F, 1948)

According to studies reported by Witzleb at the Haldane Centenary Symposium (Cunningham and Lloyd, 1963) in cats under chloralose-urethane anesthesia, which were artificially ventilated, excitation of chemoreceptors usually was apparent at an arterial Po_2 tension of about 100 mm Unless the anesthetic agents, or other conditions of the experiment sensitize the chemoreceptors to O_2 in some way, this result would indicate that influences from chemoreceptors, though slight, are concerned in regulation of respiration, even when the O_2 tension of arterial blood is within the range which occurs in a subject breathing air at sea level

CHEMORECEPTORS OF THE CAROTID AND AORTIC BODIES

The typical responses to significantly lowered oxygen tension in inspired air include *1)* stimulation of breathing, *2)* cardiac acceleration, and *3)* a rise in arterial blood pressure It is now clear that both the stimulation of breathing and vasoconstriction, which in turn causes an increase in blood pressure, can be attributed to reflex effects elicited from chemoreceptors of the carotid and aortic bodies, whereas the cardiac acceleration cannot be so explained Several researchers have demonstrated by experiments employing perfusion of the carotid bodies, that a lowering of oxygen tension in them does not cause cardiac acceleration (See M deBurgh Daly and J N Scott in Cunningham and Lloyd, 1963) Actually, cardiac inhibition occurs but much of this effect must be attributed to baroreceptor reflexes Since one of the changes elicited reflexly from the carotid bodies is vasoconstriction, the arterial blood pressure rises as a result of the increased peripheral resistance, hence if the baroreceptors of the carotid sinuses and aortic arch and their innervation are intact, reflex bradycardia is elicited However, even when the rise

in arterial blood pressure is buffered, lowering of the oxygen tension in the blood perfusing the carotid bodies does not elicit cardiac acceleration, provided the animal is also being artificially respired Daly has concluded that the cardiac acceleration which occurs in response to hypoxia is secondary to the respiratory stimulation He presents evidence that the increase in respiratory movements reflexly elicits the acceleration

From the studies of Daly, Bernthal and others, it is clear that reflex effects on heart rate from the carotid bodies cannot be determined, unless the arterial blood pressure is kept constant and artificial respiration is utilized When these conditions are met, a reduction in oxygen tension of blood perfusing the carotid body usually causes a moderate decrease in heart rate

Sodium cyanide injected intravenously causes cardiovascular and respiratory changes similar to those occurring in response to hypoxia (respiratory stimulation, increase in arterial blood pressure, cardiac acceleration) Since cyanide prevents the utilization of oxygen by tissues, it is possible that the mechanisms concerned in the response to it are the same as those involved in responses to inadequate oxygen supply Also the responses to cyanide (when it is introduced into the blood perfusing the carotid body) are similar to those produced by decreasing the oxygen tension of the perfusing blood Several other compounds have been shown to have a stimulant action on chemoreceptors of the carotid bodies, and several theories have been advanced to explain how either oxygen lack or inability to use oxygen in the presence of an oxygen excess can be transduced into impulses in the afferent fibers which innervate chemoreceptors

Recent work done in Anichov's laboratory, and other studies reviewed by him, indicate that widespread effects result from stimulation of carotid body chemoreceptors He states that the machinery of the "stress" reaction, extensively described by Selye, is set in motion, and that adrenocorticotropic hormone (ACTH), epinephrine, and norepinephrine are liberated

Ordinarily the chemoreceptors of the aortic bodies (which are much less accessible and hence have not been studied so extensively) have been considered to be of the same type as those of the carotid bodies However, Comroe and Mortimer (1964) have reported that, in dogs, hyperpnea can be elicited from the carotid bodies and aortic bodies, but from the latter arterial hypertension and tachycardia are elicited The latter effect, as explained, is not elicited from carotid bodies

EFFECTS OF HYDROGEN ION CONCENTRATION AND THE INTERACTIONS BETWEEN RESPIRATORY DRIVES

Since changes in P_{CO_2} are accompanied by changes in pH, a question may be raised whether changes in hydrogen ion concentration as such have an influence upon breathing Gray (1950) has emphasized the point that in metabolic acidosis pulmonary ventilation is greatly increased while arterial P_{CO_2} is below normal Obviously then if it is true to begin with that only three drives (P_{O_2}, P_{CO_2}, cH) operate to control breathing, increased acidity per se is capable of stimulating breathing since neither of the other two factors has changed in a direction which would cause such stimulation Gray further states that an increase in RMV is not regularly associated with an increase (and in fact may occur in the presence of a decrease) in drive related to any one of the three factors considered to be concerned with the control of pulmonary ventilation, however, at least one of three changes must be present *1)* rise in arterial P_{CO_2}, *2)* decrease in arterial P_{O_2}, or *3)* increase in arterial cH It is envisaged that any one of these may assume the major role in driving respiration under certain circumstances In the case of administering CO_2 the arterial cH increases almost linearly with the increase in arterial P_{CO_2} so that, according to Gray's multiple factor theory, the stimulation of breathing is greater than would be produced by either the rise in P_{CO_2} alone or the rise in arterial cH alone

The evidence seems to be sufficient to establish that the level of pulmonary ventilation in subjects at rest is determined by either the arterial CO_2 tension, the arterial O_2 tension, the arterial cH or by interaction of these three drives The relationship between each of these drives and the other two has been discussed at length in physiological literature Gray advanced the hypothesis that effects of the three factors are simply *additive*, and, in line with this view he presents an equation for calculation of the ventilation ratio (ratio of alveolar ventilation under new conditions to alveolar ventilation under control conditions) from the changes in P_{CO_2}, P_{O_2} and cH of arterial blood

According to Gray's theory if P_{CO_2} is lowered further below the level which leads to apnea because of lack of a CO_2 stimulus to the respiratory center, the P_{CO_2} effect is further subtracted Stated in another way, if breathing is put on an oxygen-lack drive, the lower the P_{CO_2}, the lower will be the amount of respiratory stimulation because of the lesser P_{CO_2} effect to be added to the oxygen-lack drive In experiments by Nielsen and Smith (1952) the alveolar P_{O_2} tension was maintained at constant low levels, eliciting O_2-lack drive, while the effects of adding CO_2 were tested It was found that CO_2 had no effect, or only a slight effect, on the ventilation under these conditions until a threshold value of arterial P_{CO_2} of about the same magnitude as that which causes initiation of breathing under normal conditions was reached Hence, it appears that the CO_2 level does not at all influence the level of ventilation when breathing is being maintained by the hypoxic stimulus Christensen (1954) comments that it is difficult to see how Gray's multiple factor theory can explain an unchanged ventilation when the alveolar and arterial P_{CO_2} are changed between 21 and 33 mm Hg at a constant alveolar P_{O_2} It seems possible that just as P_O is not a factor in respiratory regulation when it is above a certain level, P_{CO_2} may not be a factor at all when it is below a certain level The interaction of chemical stimuli in driving respiration is discussed in detail in the recent review by Leusen (1969) He concludes that the influences of $P_{A_{CO_2}}$ and cH are additive, whereas the combined effects of $P_{A_{CO}}$ and $P_{A_{O_2}}$ cannot be equated to an algebraic sum of the effects of the two stimuli As shown by Eyzaguirre and Lewin (1961) and Eyzaguirre and Koyano (1965) the number of impulses elicited by a given change in $P_{A_{CO}}$ or in cH in the peripheral chemoreceptors is maximal in the presence of a low local P_{O_2} but minimal when the P_{O_2} is very high, thus it appears that the effect of high P_O is virtually to block the stimulant effect of increased P_{CO_2}

Immediate respiratory responses to exposure to a given reduction in P_{O_2} in inspired air are of lesser degree than those which are seen after chronic exposure Acclimatization occurs so that during prolonged exposure pulmonary ventilation is increased, as indicated by a further lowering of the alveolar P_{CO_2} beyond that observed initially Rahn and Otis (1949) suggest that a rise in alkalinity of the blood due to decreased oxyhemoglobin saturation inhibits the respiratory stimulation that would be expected to result from the decrease in P_{O_2} per se Regardless of the mechanism of acclimatization which occurs during chronic exposure to lowered P_{O_2} in inspired air, when it has occurred, there is increased pulmonary ventilation, lowered alveolar P_{CO_2}, increased alveolar P_{O_2} beyond that observed initially, and a lowered alkali reserve During the process of acclimatization, alkali is removed by urinary execretion The pH of the blood is within the normal range It appears that other more complex changes in respiratory control are involved in the gradual adjustments to hypoxia For example, the alveolar P_{CO_2} which is lowered initially may return to or even exceed the level in normal subjects at sea level, hence sensitivity of the regulatory mechanism to CO_2 must have become readjusted In short, in acute hypoxia the CO_2 drive may be largely or completely eliminated, but it must be assumed that the CO_2 drive is in some degree restored during acclimatization

Other Patterns of Ventilation

MECHANISMS OF HYPERPNEA DURING EXERCISE

During exercise pulmonary ventilation increases, and the rise is proportionate to the amount of increase in muscular activity Since an increased rate of production of CO_2 and of acid metabolites and an increased rate of utilization of oxygen occur as a consequence of increased muscular activity, a priori it is logical to think that the respiratory stimulation is induced by the same three factors which are considered to operate in the resting subject However, a number of studies show that muscular exercise is not regularly associated with either a decrease in arterial P_{O_2}, or a rise in arterial P_{CO_2} or cH In individual experiments none of these factors shows a change during exercise in the direction needed to elicit increased respiratory activity (Grodins, 1950)

The P_{CO_2} of *mixed venous* blood is increased and the P_{O_2} decreased during exercise, but no respiration-influencing receptor located on the *venous* side of the systemic circuit has been demonstrated (The presence of bodies similar to the aortic bodies around the pulmonary artery and receiving a blood supply from it have been described, and if such receptors exist they might be expected to elicit stimulation of breathing during exercise) There is evidence that respiration is stimulated reflexly by im-

pulses arising in the working muscles or the joints which are acted upon by these muscles Still another possibility is that during voluntary activity impulses are directed in descending nerve tracts into the respiratory center to exert a stimulant effect, but there appears to be no evidence that this is the mechanism (see Kao in Cunningham and Lloyd, 1963) In any case it appears that accessory respiratory drives are brought into play during exercise so that increased ventilation may occur even before a change in P_{CO_2}, P_{O_2} or cH have had time to occur at the sites upon which these stimuli act The factors which are concerned in the stimulation of breathing, which most respiratory physiologists believe must be present in exercise over and above that related to the three chemical stimuli which operate under resting conditions, are referred to by Grodins (1950) as the *exercise stimulus* It is probable that afferent influences from various sources are concerned and also that chemical excitants not yet identified play a role

PERIODIC BREATHING

Under certain conditions irregular or "periodic" breathing may develop There are two major types Biot's breathing and Cheyne-Stokes breathing In Biot's breathing there are periods of variable length during which breathing is present, and these alternate with apneic periods of different lengths The transitions from the periods of activity to apnea and back are abrupt In contrast to this in Cheyne-Stokes breathing, although periods of respiratory activity alternate with apneic periods, the amplitude of breathing increases gradually to a maximum and then decreases gradually until the next period of inactivity Biot's breathing has been observed in meningitis and other pathological conditions affecting the medulla Cheyne-Stokes respiration has been observed in a number of different circumstances, some of which are pathological and some physiological (Fig 6 31)

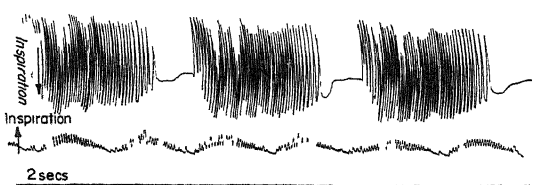


Fig. 6 31 Two examples of periodic breathing (After Lewis)

Cheyne-Stokes breathing occurs in some normal subjects at moderately high altitudes, (for example, 12,000 to 15,000 ft) and it may be seen in normal sleeping infants Experimentally, it may be produced in normal subjects following a prolonged period of voluntary hyperventilation, and it can be produced by having the subject breathe deeply through a tube of a certain length In the latter case, the dead space is increased by a specific amount, and if there is too little or too much increase in dead space, Cheyne-Stokes breathing will not develop Several drugs, notably morphine, may cause the appearance of Cheyne-Stokes breathing The condition is commonly seen in persons suffering from chronic congestive heart failure, and in such cases the periodicity tends to be aggravated by sedation or during sleep

The basis for Cheyne-Stokes breathing after hyperventilation seems to be that oxygen-lack drive appears and then is relieved several times before the P_{CO_2} finally is built back up to levels sufficient to drive normal respiration (see above) The oxygen-lack drive presumably is elicited reflexly from peripheral chemoreceptors However, there is evidence that two or more types of initial defect may cause Cheyne-Stokes breathing For example, during periodic breathing there are also cyclic changes in the blood pressure level, and in some instances the rise in mean arterial blood pressure occurs during the phase of respiratory activity while in other cases the rise in blood pressure is during the apneic phase These two types most likely are related to quite different basic mechanisms The former has been demonstrated to occur in dogs having the sino-aortic baroreceptors and the carotid and aortic bodies denervated or isolated (Youmans and Schopp, 1957) hence, chemoreceptive drive could not have been a factor It is still probable, however, that the other type of periodicity is dependent upon a seesawing between reflex and central drives

Guyton et al (1956) have described experiments in which Cheyne-Stokes breathing was produced if the time required for blood to flow from the heart to the brain was prolonged by increasing the length of the channels Since circulation time is increased in chronic heart failure, this may be the change that leads to periodic breathing in patients suffering from this condition

Eyster (1906) produced periodic breathing in dogs by elevating the intracranial pressure His records show that breathing occurred when

blood pressure rose above the level of the intracranial pressure, and apnea was present when the pressure was below this level In clinical cases with increased intracranial pressure, Eyster observed the same pattern of cyclic changes in respiration and blood pressure as that produced experimentally in dogs

DYSPNEA

Dyspnea literally means difficult breathing When the respirations cannot be carried out with ease and practically unconsciously, the individual is said to be dyspneic The term therefore implies a subjective element Dyspnea thus differs from hyperpnea The latter term means simply increased pulmonary ventilation, and this may occur quite unconsciously or if the subject is aware of the augmentation of the breathing, there is not necessarily any sensation of difficulty or distress When the hyperpnea becomes extreme and yet leaves the need for which it has been instituted unsatisfied, discomfort or distress is experienced, and the term dyspnea is applicable Meakins offers the following concise definition "Dyspnea is the consciousness of the necessity for increased respiratory effort "

Since the respiratory and circulatory functions are directed toward the acquisition of oxygen and the elimination of carbon dioxide, dyspnea may result if either of these functions be disturbed to such an extent that the normal gaseous exchanges cannot be accomplished On the other hand, the oxygen requirement and the carbon dioxide production may be so great that the *normal* respiratory and circulatory mechanisms find difficulty in meeting the demands of the moment Or again the supply of oxygen itself may be inadequate as a result of a low oxygen tension in the atmosphere (e g , high altitudes) In considering the causes of dyspnea we must, therefore, in many instances look beyond the lungs themselves

The fundamental or immediate causes of dyspnea can in most instances be reduced to the following categories *1*) Stimulation of the respiratory center either *a*) reflexly from the peripheral chemoreceptors (anoxia or increase in H-ion concentration due to fixed acids), *b*) directly, by CO_2 excess, *c*) a combination of *a* and *b* as in asphyxial states, or *d*) by impulses from cerebral centers or by afferent impulses, especially of a painful character, from abdominal or peripheral regions *2*) Hypersensitivity of the Hering-Breuer reflex, thus bringing about earlier inhibition of the inspiratory phase and causing, as a consequence, a more rapid shallow type of breathing

The abnormal conditions chiefly associated with dyspnea are

1) Prevention of adequate oxygenation of the blood in the lungs (arterial hypoxia, chap 5)

2) Interference with the transport of the respiratory gases Anoxia of the hypokinetic or anemic type *a*) slowing of the circulation as in cardiac failure, *b*) severe anemia, breathlessness chiefly on exertion

3) Restriction of the action of the diaphragm or intercostals

4) *Acidosis*, reduced alkali reserve or retention of CO_2

5) Increased metabolism

6) *Nervous conditions*, e g , emotional disturbance, neurasthenia, hysteria, encephalitis, or the direct stimulation of the respiratory center by cerebral tumor, hemorrhage or edema

Dyspnea Due to Pulmonary Diseases. Dyspnea is a feature of various respiratory diseases as follows

1) In some instances, e g , laryngeal or bronchial obstruction and asthma, the dyspnea is due to a combination of anoxia and CO_2 retention

2) In other instances owing to the reduced distensibility of the lungs resulting from edema, congestion, inflammation, fibrosis, etc the Hering-Breuer reflex is thought to be abnormally sensitive

3) Limitation of the movements of the diaphragm and chest wall In emphysema, for example, owing to the loss of lung elasticity the resting position of the chest is one of nearly full inspiration The diaphragm is fixed and the thorax elevated Any further enlargement of the chest entails unusual effort on the part of the intercostal muscles (see also chap 3) and the enlistment of the accessory muscles of respiration Expiration involves active contraction of the expiratory muscles

Cardiac Dyspnea. Dyspnea upon exertion is a symptom in patients having certain chronic pulmonary and heart lesions, e g , mitral stenosis Stimulation of the carotid and aortic bodies by oxygen want or of the respiratory center by carbon dioxide excess probably is not in most cases responsible for the dyspnea, since the oxygen saturation of the arterial blood may not be reduced to any important degree, and the

carbon dioxide tension is within or even below the normal range, and cardiac output may be normal Pulmonary engorgement leading to diminished distensibility of the lung is considered to be the prime cause of cardiac dyspnea Although the reduction in the vital capacity may be roughly proportional to the dyspnea, the two do not bear the relationship of cause and effect Due, however, to the diminished distensibility—the stiffness of the lung—a greater inspiratory effort is expended in breathing the extra volume of air which the muscular exertion demands The recoil of the lung is also moderately reduced so that expiration instead of being a passive act may require the aid of the contraction of expiratory muscles The intrapleural pressure, therefore, instead of remaining "negative" throughout the respiratory cycle may become positive toward the end of expiration (Christie and Meakins, 1934) The reduced distensibility possibly also has the effect, as already mentioned, of increasing the sensitivity of the Hering-Breuer reflex with the production of shallow breathing

When pulmonary edema supervenes in patients with congestive heart failure, interference with the uptake of oxygen and the production of arterial hypoxia (stimulation of carotid body) may possibly be a factor Hindrance to the absorption of oxygen, caused by the presence of exudate in the alveoli and the edematous swelling of the alveolar walls, is accompanied by little or no interference with the elimination of carbon dioxide, this is probably due in part to the much greater solubility of carbon dioxide than of oxygen and, in consequence, to the freer diffusion of the former gas through the edema fluid In congestive heart failure arterial hypoxia with a normal or even a subnormal carbon dioxide content of the arterial blood may exist, even in the absence of pulmonary edema

Experimental support can be cited for the view that reduced distensibility of the lungs as a result of congestion is an important factor in cardiac dyspnea It has been mentioned elsewhere (Chap 6) that distension of the pulmonary bed causes rapid shallow breathing

Though there is much to be said for the reflex origin of cardiac dyspnea, not all are agreed as to its paramount importance. Another possibility is that the hyperpnea and dyspnea of the cardiac patient *during rest* are due to reduced blood flow through the respiratory center, resulting in a high CO_2 tension, and possibly to the accumulation of acid products of its own metabolism

Orthopnea In congestive heart failure with dyspnea at rest the breathlessness is usually more pronounced in the recumbent than in the sitting position When propped up with pillows, the patient may be comfortable but becomes dyspneic when he lies down Many theories have been advanced in attempts to explain the less difficult breathing in the upright position Among these are

a) removal of the weight of the abdominal viscera which interferes with the descent of the diaphragm in the recumbent position, *b*) reduction in the intracranial venous pressure and the improved draining of blood from the medulla and in consequence, augmentation of the flow through the respiratory center, *c*) draining of blood from the pulmonary circuit This is probably the most important factor The vital capacity is less in the recumbent than in the sitting posture This is true even for the normal person, but in cardiac cases the effect is accentuated by the pulmonary vascular engorgement and decreased distensibility of the lung induced by recumbency

Cardiac Asthma (Paroxysmal Nocturnal Dyspnea) In some patients with severe chronic congestive heart failure acute dyspnea which may terminate in pulmonary edema comes on during the night The patient has a feeling of suffocation He assumes the upright position and may go to the window for air These attacks are related to excessive engorgement of the lungs with blood so that pulmonary capillary pressure rises to the point of producing transudation into the alveoli As indicated above, the volume of blood in the lungs increases when one changes from the upright to the recumbent position Patients with chronic heart failure already have an elevated venous pressure which increases further when they recline The key question is with regard to the fact that paroxysmal nocturnal dyspnea comes on after a delay usually of several hours since the effects of changing from the upright to the recumbent position insofar as they are related to gravity might be expected to occur within seconds or minutes. The clue to this problem probably is provided in the observation that edema fluid in patients with congestive heart failure is transferred from the interstitial spaces into the blood, and a rise in blood volume occurs slowly during the night. This leads to a further elevation of venous and pulmonary cap-

illary pressure The paroxysm could be set off in two ways The pressure in the pulmonary capillaries could rise to the point of producing transudation and hence impairment of exchange of O_2 and CO_2 On the other hand, the venous pressure might reach the critical level at which overdistention of the heart would result in a decrease in the output of the heart Once this occurs a vicious cycle is produced decreased output leads to further elevation of venous pressure, and elevation of venous pressure leads to further overdistention of the heart and decreased output In either case the change to the orthostatic position would be beneficial in that the blood would tend to remain in the dependent part of the body and thus decrease the volume and pressure of blood in the heart and lungs Also, it should be noted that acute cardiac decompensation, produced by exceeding the critical pressure on the curve which relates cardiac output to venous pressure, leads to a rise in pulmonary capillary pressure and hence to acute pulmonary edema

Anemia. The anemic subject at rest typically is not dyspneic The hemoglobin, though reduced in amount, becomes saturated with oxygen in the lungs The oxygen tension and consequently the quantity of the gas in simple solution in the arterial blood are normal The arterial blood of a patient whose hemoglobin is 30% of the normal value will, however, contain only a little over 6 volumes % of oxygen In the healthy resting body the blood in its passage through the capillaries gives up about 5 volumes % If the velocity of blood flow through the tissues in anemia were the same as during health, a unit volume of blood would tend to give up an equivalent amount of oxygen If this occurred there would be a reserve of only 1 volume %, i e , the venous blood would be almost completely reduced, and the tissues including the carotid and aortic bodies, would suffer at all times from anoxia The circulation rate (cardiac output) is, however, increased in severe anemia so that each unit of blood gives up a smaller proportion than that of its oxygen load The chemoreceptors apparently are not stimulated, and the patient is not dyspneic while resting, even though his hemoglobin is greatly reduced Not only is the output of the heart increased, but a redistribution of the blood flow occurs The vessels of the skin are constricted, and a greater proportion of the total blood volume is driven through other regions The extent to which the circulatory readjustments can compensate for the blood defect is limited, therefore during exertion the extra demand for oxygen cannot be met Oxygen want follows, the respiratory mechanism is stimulated, hyperpnea and dyspnea result

It is to be remembered that in anemia the carriage of carbon dioxide may also be interfered with since hemoglobin plays an important role in the transport of this gas (see sect 4, chap 1)

Increased Metabolism. Exercise may bring about a great increase of metabolism Reflexes initiated in the active muscles and possibly the irradiation of impulses from the motor cortex are thought to play an important part in the hyperpnea of exercise (p 6–57) As the severity of the exercise is increased, hyperpnea merges into dyspnea The athlete and the untrained person differ widely in respect to the degree of muscular exertion which will produce this physiological type of dyspnea The difference depends upon the following factors

Vital Capacity In the average normal man the pulmonary ventilation increases from four to five-fold before the dyspneic point is reached The athlete on the other hand, since his vital capacity is greater, shows a correspondingly greater increase in his pulmonary ventilation before dyspnea supervenes The existence of any pulmonary condition which reduces the vital capacity will depress the level of the dyspneic point

Circulation Rate The trained man's circulation rate increases to a greater degree than the untrained, and so more oxygen is delivered to his tissues

Neuromuscular Integration Coordination of the several muscles in a given muscular act is more precise in the trained than in the untrained man There is thus less waste of effort In the performance of a given amount of work, therefore, the untrained man consumes a greater volume of oxygen, i e , he is a less efficient machine

A pathological increase in metabolism, e g , hyperthyroidism, fever, etc , does not, in the absence of some respiratory or circulatory abnormality, cause dyspnea during rest The increase in metabolism would need to be around 300% before dyspnea might be expected to occur and such a metabolic level is never reached in these or any other pathological condition Nevertheless, in hyperthyroidism, dyspnea will follow upon a degree of exertion which would cause no distress in a healthy per-

son, for the greater metabolic rate, due to the disease, added to that of the exercise will increase the metabolism sufficiently to raise the pulmonary ventilation above the dyspneic point The dyspnea of hyperthyroidism, when pulmonary and cardiovascular complications are absent, therefore, is like that of anemia, in that it is evident only upon exertion

Acidosis. The importance of pulmonary ventilation in resisting a rise in the hydrogen ion concentration of the body fluids has been dealt with elsewhere (Sect 4, chap 11) Nonvolatile acids, e g , lactic in muscular exercise, β-hydroxybutyric and acetoacetic acids in diabetes, and retained acids in nephritis, react with bicarbonate The alkali reserve becomes reduced, but the ratio,

$$\frac{H_2CO_2}{NaHCO_3} = \frac{1}{20}$$

is maintained through the stimulating effect of CO_2 upon the respiratory center It is thus that the CO_2 of the alveolar air (and so of the arterial blood) is kept at a level proportional to the reduction in the denominator of the equation When this can no longer be effected, i e , when the hydrogen ion concentration of the blood rises, the acidosis being then uncompensated, the center is stimulated both directly and through reflexes initiated from peripheral chemoreceptors

The production of fixed acids in diseased conditions rarely causes such a high degree of hyperpnea that dyspnea results, unless the circulatory and respiratory mechanisms are very inefficient or the metabolic rate is increased According to Means the "alkali reserve" must drop to around 12 volumes % before dyspnea supervenes as a result of the acidosis itself In milder grades of acidosis, however, dyspnea occurs upon exertion The hyperpnea due to the increased metabolism of the exercise is then added to that due to the acidosis, with the result that the dyspneic point is soon reached

In severe diabetic acidosis *air hunger*, or Kussmaul breathing, may be seen in the comatose patient

Carbon dioxide retention may, as in emphysema, be accompanied by a compensatory rise in bicarbonate and a normal blood reaction (compensated CO_2 excess), dyspnea is not a notable feature in this condition In other instances of CO_2 retention, compensation is incomplete (uncompensated CO_2 excess or gaseous acidosis) and dyspnea results In others, again, the CO_2 retention is due to the depression of the respiratory center itself, as in morphine narcosis, in such instances, although compensation is incomplete, dyspnea of course does not occur

5 / HYPOXIA, ASPHYXIA, DYSBARISM, OXYGEN THERAPY, AND RESUSCITATION

Types of Hypoxia

The terms *hypoxia* and *anoxia* are synonymous, although the former etymologically is more suitable Both terms are in common use and will be used interchangeably in the following discussion The most useful definition of hypoxia is that it refers to any condition in which there is an inadequate *supply* of oxygen to the tissues According to this definition it is synonymous with *oxygen lack* or *oxygen want* Upon the basis of three quite different categories of causes Barcroft classified anoxia into *anoxic*, *anemic*, and *stagnant* types The categories described by him still need to be differentiated, but changes in the names of two of them seem to be preferable A fourth type, caused by cyanide (p 6-78) and known as histotoxic anoxia, was added by Peters and Van Slyke, however, there is no oxygen lack in this case, but rather inability of the cells to utilize the oxygen brought to them, with the result that oxygen tension of the tissues is actually elevated Hence, it would seem to be preferable not to classify cyanide poisoning as a form of hypoxia For reasons which will become apparent in the subsequent discussion *arterial hypoxia* is preferable to anoxic anoxia in Barcroft's classification, and *hypokinetic hypoxia* is a better term for stagnant anoxia The three major types of oxygen lack, together with the principal causes of each are listed below Also, it should be realized that two or more of these types or causes can be present simultaneously, and each type may be either acute or chronic

OXYGEN LACK
MAJOR TYPES AND CAUSES

Arterial Hypoxia. Arterial hypoxia *is characterized by a lower than normal* Po_2 *in arterial blood* (Oxygen capacity of blood and rate of blood flow are normal or elevated)

A Low Po_2 in inspired air
  1 Altitude
  2 Breathing in closed space[1]
  3 Breathing artificial gas mixture containing lower Po_2 than that found in atmospheric air

B Decreased pulmonary ventilation
  1 Airway obstruction[1]
  2 Weakness or paralysis of respiratory muscles[1]
  3 Depression of respiratory center by drugs[1]
  4 Pneumothorax

C Inadequate oxygenation in (abnormal) lungs
  1 Poor mixing[1]
  2 Impaired diffusion (alveolar-capillary block)
  3 Bronchiolar constriction (asthma)[1]
  4 Filling of alveoli with fluid Pulmonary edema, pneumonia, pulmonary hemorrhage, drowning[1]

D Venous-arterial shunts Various types of cardiac or vascular abnormalities in which unoxygenated blood is mixed with oxygenated blood

Anemic Hypoxia. Anemic hypoxia *is characterized by a lowered oxygen capacity of the blood* (The Po_2 of arterial blood and rate of blood flow are normal or elevated)

A Less than normal hemoglobin content *Anemias* of all types
B Hemoglobin combined with something other than oxygen, e g , *carbon monoxide*
C Hemoglobin altered so that it cannot combine with O_2 *Methemoglobin* found after poisoning with chlorates, nitrites, ferricyanides, acetanilid, etc

Hypokinetic Hypoxia. Hypokinetic hypoxia *is characterized by a decreased rate of blood flow*

A Generalized Congestive heart failure, hemorrhage, shock, etc
B Localized Vasospasm, thrombosis, embolus

[1] Arterial Pco_2 is increased

Some confusion has arisen in the use of the term *anoxemia* In some of the medical dictionaries and textbooks it is defined as synonymous with arterial hypoxia while some authors (Comroe and Dripps, 1950) use it as referring to a diminution of oxygen in the blood whether it be related to decreased Po_2 or decreased hemoglobin

Comroe (1965) and Lambertsen (1968) believe that the designation of another type of tissue anoxia as overutilization anoxia is needed This is to refer to a decrease in Po_2 in tissues related to increased utilization of oxygen by the tissue, as in severe muscular exercise Also Lambertsen (1968) suggests the name "hypotonic anoxemia" for arterial anoxia and "isotonic anoxemia" for anemic anoxia

Asphyxia. *Asphyxia* refers to conditions in which hypoxia is combined with *hypercapnia*—an increased CO_2 tension in the arterial blood and hence in the tissues also It will be noted that a number of the causes of arterial hypoxia also lead to hypercapnia, while in some cases, notably at altitude, hypercapnia is not a factor The opposite change, a lowered CO_2 in the body, is referred to as *hypocapnia* or *acapnia* The reactions in asphyxia are related to effects of decreased arterial and tissue Po_2 and increased arterial and tissue Pco_2 Asphyxia of some degree is present in a number of those conditions listed in the classification of causes of oxygen lack, since elimination of CO_2, as well as uptake of oxygen, is impaired Respiratory stimulation and dyspnea are prominent, and the stimulation is derived from both the lowered Po_2 and increased Pco_2 Cyanosis, being related to the presence of an increased amount of reduced hemoglobin in the blood in the cutaneous capillaries (p 6–78), also is characteristic of asphyxial states as well as those in which arterial anoxia is present without hypercapnia Blood pressure tends to be elevated due to the stimulant action of CO_2 on the vasoconstrictor center, except in the terminal stages Chronic adjustments to asphyxia include the increase in erythrocyte count, hematocrit, hemoglobin, and oxygen capacity of the blood, which are induced by the lowered Po_2 tension, and renal excretion of acid to compensate for the rise in cH produced by retention of CO_2

TENSIONS AND VOLUMES PER CENT OF OXYGEN IN THE BLOOD IN HYPOXIA

In the *arterial* type of hypoxia both the Po_2 and the volumes per cent of O_2 are lowered The blood has a normal or elevated O_2 capacity (20 volumes % or more), but it is not exposed to Po_2 which is high enough to produce the usual degree of saturation (95 + %) However, even though the blood is carrying less than a full load of O_2, when it reaches the capillaries in tissues which are functioning at a lower than normal oxygen tension it may unload as much oxygen/100 ml of blood as occurs normally That is, normally about 20 volumes % would be present in the arterial blood and 15 volumes % in mixed venous blood, while in arterial hypoxia arterial blood capable of holding 20 volumes % might actually contain 15 and give off 5 so that the mixed venous blood would contain only 10 volumes %. In the latter case 50% desaturation of hemoglobin would be produced at the tissue level as compared with 25% in the normal

In *hypokinetic* hypoxia both the Po_2 and volumes per cent of O_2 in the arterial blood are normal, however, as the blood flow through the tissues is slower than in the normal state and the cells continue to use oxygen at an essentially normal rate, the Po_2 in the tissues becomes lowered, and this causes the unloading of more O_2 per unit volume of blood flowing through the tissues In very active tissues which are subjected to a severely deficient blood flow, virtually all of the oxygen in the blood delivered to the tissue may be extracted This indicates that the Po_2 in such tissues approaches zero Such a situation can not exist for more than a very short time before damage to the tissue results

Hypokinetic hypoxia may be generalized as in conditions in which the cardiac output is significantly reduced below the level which would be present in a normal subject under similar circumstances On the other hand a localized reduction in blood flow may occur in specific tissues or organs and lead to reduction in or losses of functions of the involved organ Myocardial infarction, however, is an example of localized hypoxia resulting in loss of a function which leads to generalized hypoxia and worsening of the hypoxia in the myocardium

In *anemic* hypoxia the arterial Po_2 is normal so that saturation of hemoglobin approaches 100%, but the volumes per cent of O_2 carried is reduced in proportion to the reduction in the amount of hemoglobin Hence, if blood carrying, for example, 10 volumes % of O_2 at full saturation of the hemoglobin gives up 5 volumes %/100 ml of blood flowing through the

tissues, it is apparent that the O_2 in the tissues is so low as to cause 50% desaturation Again this is to be contrasted with the normal state in which, on the average, only about 25% desaturation occurs as blood flows through the tissues

Arterial Hypoxia at Altitude

Arterial hypoxia, uncomplicated by hypercapnia, is observed in individuals breathing air at high altitudes or in decompression chambers Knowledge of both the acute and chronic adjustments to altitude is of great interest since man ascends rapidly in airplanes or resides for years at levels of 15,000 to 18,000 ft

The per cent composition of atmospheric air is the same at altitude as at sea level But the barometric pressure decreases with altitude It is noteworthy that at an altitude of 18,000 ft the barometric pressure is about 380 mm Hg or one-half of that at sea level The oxygen tension, therefore, is diminished to one-half of that at sea level, since the Po_2 at any altitude is calculated by multiplying the barometric pressure by 0 2094 Thus, a second vertical scale showing oxygen tension could be derived simply by multiplying each of the figures in the barometric pressure scale by 0 2094

At altitude as one breathes air containing a lower Po_2 than that of air at sea level the alveolar Po_2 becomes lowered The decrease in alveolar Po_2 at altitude if severe enough is counteracted in part by an increase in pulmonary ventilation, although this response is not as prominent immediately as it is after some delay

Conditions at altitude can be simulated in an air tight chamber from which the air has been partially evacuated At a given barometric pressure a certain altitude is simulated as far as gas tensions are concerned Chambers have been constructed in such a way that they may be very rapidly, i e , explosively, decompressed and the reactions of the subject followed from moment to moment. During acute exposure to simulated altitude, produced by rapid decompression of a chamber containing the subject, alveolar ventilation is increased so that alveolar Pco_2 is lowered, and the amount of decrease is proportional to the reduction in ambient pressure (increase in altitude) For example, shortly after decompression to a simulated altitude of 18,000 ft the alveolar Pco_2 is in the range of 27 $\pm$ 2 mm Hg, and the alveolar Po_2 is in the range of 42 $\pm$ 3 mm Hg The alveolar Po_2 is less than half that found at sea level At this point it should be noted that the water vapor tension in alveolar air, being determined by body temperature which is constant, remains at 47 mm Hg at all altitudes, and when expressed in terms of percentage, it increases in direct proportion to the decrease in total pressure in the alveoli For example, at sea level the percentage of water vapor is $47/760 \times 100$ whereas at 18,000 ft it is $47/380 \times 100$ At an altitude where the barometric pressure is only 74 mm Hg, water boils at body temperature and the alveoli are filled with water vapor

Immediately after sudden exposure to altitude, respiratory adjustments occur, however, conditions continue to change from one 5-min period to the next for a period of 30 to 60 min (Riley, Otis, and Houston, 1954) Some of the bases for these changes already have been discussed

The immediate circulatory adjustments to acute arterial hypoxia are an increase in heart rate and blood pressure, the cyanosis that appears is a result of an increase in the amount of reduced hemoglobin in arterial blood The amount of desaturation can be estimated from the alveolar Po_2 and the oxygen-hemoglobin dissociation curve

In the normal subject breathing air (about 21% O_2) about 95% saturation of the hemoglobin of arterial blood with oxygen is attained In a person breathing air at any given altitude somewhat less than 95% saturation would occur but might be achieved if the O_2 percentage in inspired air were increased sufficiently It is found that at 18,000 ft it would be necessary to breathe about 45% O_2 in order to attain 95% saturation, and at about 34,000 ft , 100% O_2 is needed It follows that at altitudes higher than 34,000 ft some degree of arterial hypoxia will occur even when the subject breathes pure O_2, and at very high altitudes intolerable degrees of oxygen lack would be experienced even when breathing 100% O_2 The limit at which aviators can fly in nonpressurized cabins breathing 100% O_2 is in the range of 40,000 to 45,000 ft Above this level cabins containing gases under pressures higher than that of the ambient air are required

Several mountain climbing expeditions have been made in the past by different groups of physiologists for the purpose of studying the effects of low oxygen tensions upon the respiratory functions and of determining the factors underlying the phenomenon of acclimatization An expedition was made to Monte

Rosa (15,000 ft) in 1894 by Mosso and by others subsequently, to the peak of Teneriffe (12,000 ft) in 1910 by Zuntz, Barcroft, and associates, and in 1911 to Pike's Peak (14,100 ft) by the Anglo-American expeditions of which Haldane, Douglas, Henderson, and Schneider were members In 1921 to 1922 Barcroft headed a party to Cerro de Pasco (14,200 ft) in the Peruvian Andes A Himalayan expedition was led by Hartman in 1931, and in 1935 Dill led a party to the Chilean Andes Pugh (1957) reported observations made on Mt Everest during the successful British expedition of 1952

GENERAL SIGNS AND SYMPTOMS

Airplane ascents, if made rapidly without the use of oxygen, may result in sudden loss of consciousness due to the reduction in oxygen supply to the brain When the ascent is made more slowly, or the altitude is not so great as to cause immediate loss of consciousness, the aviator may at first experience sensations of excitement, exhilaration and well being As higher altitudes are reached effects of a more serious nature develop, often insidiously Mental and sensory dullness, muscular weakness, headache, vomiting, cyanosis, dyspnea, and perhaps a tendency toward periodic breathing may be induced A common and dangerous effect is the development of fixed ideas which may result in the performance of foolhardy and ill-judged actions

When a person climbs to a mountain height, the time taken in the journey allows a certain degree of physiological readjustment to take place, and the symptoms are usually less intense But, as in the case of the aviator, mental features, e g , a feeling of elation, exhilaration, talkativeness, and sometimes emotional outbursts, laughing or crying, quarrelsomeness, or the development of fixed ideas are prominent Mental tasks, e g , calculations, memory tests, and telling the time from the mirror image of a clock face are performed less efficiently Similar effects upon the mind are produced upon persons exposed to low oxygen pressures within a steel cabinet The mental effects as pointed out by Barcroft are not unlike those caused by drunkenness To quote his words,

"Alcohol affects different persons in different ways, so on my journeyings in high altitudes I have seen most of the symptoms of alcoholism reproduced I have seen men vomit, I have seen them quarrel, I have seen them become reckless, I have seen them become morose I have seen one of the most disciplined of men fling his arms about on the ledge of a crevasse to the great embarassment of the guide I have seen the most loyal companion become ill-tempered and abusive to the point at which I feared international complications would arise "—Lessons from High Altitudes

Complete prostration may follow the earlier symptoms If the individual remains at the high altitude the symptoms pass off after a time, as he becomes acclimatized to the low oxygen tension

At a simulated altitude of about 20,000 ft (i e , in a decompression chamber) most subjects experience failing vision, incoordinations, and inability to write or to execute simple mental tasks within 15 min or so after the extra oxygen supply has been cut off, at 26,000 ft none are able to retain "effective consciousness" for this length of time without oxygen

CHANGES ASSOCIATED WITH ACCLIMATIZATION

At an altitude of 14,200 ft as at Cerro de Pasco where Barcroft and his party carried out their investigations, the barometric pressure is around 440 mm Hg The partial pressure of oxygen is therefore about 92 mm Hg The oxygen tension of the alveolar air is not as far below that of the atmosphere as at sea level, and varied among the greater number of the party from 55 to 60 mm Hg The closer approximation of the atmospheric and alveolar oxygen tensions at high altitudes is due to the increased breathing which is brought about through the action of the lowered oxygen tension in the blood The oxygen tension is slightly lower in the arterial blood than in the alveolar air which indicates that the passage of the gas is due purely to diffusion and not to an active secretion by the pulmonary epithelium If this occurred, as had been suggested (Haldane), one would expect the arterial oxygen tension to be higher than that of the alveolar air Alveolar carbon dioxide tension also, as a result of the increased pulmonary ventilation, is lower than that at sea level, it varied in different individuals of Barcroft's party from 23 to 29 mm Hg At 14,200 ft the arterial blood is from 85 to 88% saturated with oxygen

A marked increase in the number of red cells (see sect 4, chap 3) and a corresponding increase in hemoglobin content of the blood occur at high altitudes The blood volume is also augmented The natives of mountainous regions

have a red cell count of from 6 to 8 million/mm^3 The greater quantity of hemoglobin of course raises the oxygen capacity of the blood and so tends to compensate for its lowered oxygen saturation That is, the *total* oxygen content of the arterial blood tends in spite of the low saturation to rise Nevertheless, it may not be evident at first sight how a rise in the oxygen capacity is of advantage, for blood of normal hemoglobin content even when only 80% saturated possesses a quantity of oxygen which is quite adequate for the needs of the tissues It has already been pointed out, however, that the important factor in supplying the tissues is the oxygen pressure gradient between the plasma in the capillaries and the tissue cells So then, if there are a greater number of red cells, each will be required to give up less of its oxygen store in passing through the capillaries to furnish a given quantity of oxygen (see anoxia due to anemia, p 6-80) Consequently, the saturation and the oxygen tension of the venous blood will be maintained at a higher level than otherwise would be possible This means that the mean intracapillary oxygen pressure will also be higher, and as a result the tissues are more effectively supplied with oxygen

The reduction in the alveolar carbon dioxide results in a corresponding decrease in the carbon dioxide tension of the arterial blood The ratio $H_2CO_3/NaHCO_3$, which tends to be altered by the loss of carbon dioxide, is adjusted by a decrease in the excretion of acid and ammonia in the urine, a consequent lowering of the "alkali reserve" and depression of the CO_2 dissociation curve (p 6-15) The actual pH of the plasma changes little if at all Up to about 12,000 ft if any change occurs it is toward the alkaline side, above this level the blood reaction shows little further change or tends to return to normal Lactic acid, which was thought at one time to be produced in excess as a result of the anoxia, is actually formed in smaller amounts at high altitudes than at sea level Even during severe exercise at 15,000 ft and higher altitudes, the lactic acid concentration in the blood is lower than during exercise of comparable severity at sea level

Barcroft and his party observed a shift to the left in the oxygen dissociation curve of hemoglobin, i e , the affinity of hemoglobin for oxygen was increased The shift in the dissociation curve is ascribed by Barcroft to an increased alkalinity of the *interior of the red cell* This increased alkalinity is in turn a direct result of the rise in the number of red cells The buffering power of the blood is increased through the greater facility offered for the action of the "chloride shift" mechanism (pp 6-11 and 6-18) In other words when a given amount of carbon dioxide is liberated by the tissues, it is distributed among a greater number of red cells than under normal circumstances, therefore, the alkalinity of each cell is reduced to a proportionately less extent There have been conflicting reports concerning this question of the shift in the oxygen dissociation curve Some observers have been unable to confirm Barcroft's finding, while others claim that a shift to the right occurs It appears that up to about 14,000 ft the affinity of hemoglobin for oxygen increases, but at higher levels the dissociation curve tends to assume the form found at sea level, and at altitudes of 19,000 ft there is a definite shift to the right

One might suppose that an increased circulation rate would be an important adjustment to the rarefied atmosphere, whereby an adequate oxygen supply to the tissues would be maintained, but except for a temporary increase during the first few days, no change in cardiac output occurs at altitudes of less than 14,000 or 15,000 ft Above 15,000 ft the greater degree of anoxia results in an increase in the minute output of the heart Before acclimatization, the pulse rate during rest increases by from 15 to 20 beats/min at altitudes between 15,000 and 18,000 ft At greater altitudes, especially in persons in poor physical condition, the increase in rate may be greater than in the normal The blood pressure shows little or no change up to 15,000 ft , a small rise may occur at higher altitudes

Those who have lived all their lives at very high altitudes (around 14,000 ft) have a larger vital capacity than dwellers at sea level Moderate altitudes, up to about 7500 ft , appear to have little or no effect upon the chest development

Although anoxia is the most serious effect of high altitudes with which the mountain climber or aviator has to contend, rapid ascents, as by airplane, cause other important physiological disturbances which should be mentioned, namely *1*) expansion of gases in the gastrointestinal tract, *2*) aeroembolism (see below), and *3*) pressure disturbances in the ears

Expansion of Gases in the Gastrointestinal Tract Like the gases of the atmosphere those in the stomach and intestine increase in volume in proportion to the reduction in pressure Gas having a volume of 1 liter at sea level expands to 2 liters at the pressure (375 mm Hg) existing at 18,000 ft , to 4 liters at 34,000 ft , to 6 liters at 42,000 ft (pressure 128 mm Hg) Distention of stomach and intestine will

result unless the abdomen is supported by a belt or other means, or the gases are freely evacuated In rapid ascents distress or even severe pain results if there is any hindrance, as by an obstruction in the colon, to the ready passage of flatus

Dysbarism

The term *dysbarism*, or decompression sickness (see Behnke, 1965), refers to the signs and symptoms which result when the total pressure of ambient gases suddenly is reduced It refers to the effects of reduction in pressure as such Dysbarism is seen in deep sea divers or caisson workers when they pass too quickly from the high pressure in which they have been working to the ordinary pressure of the atmosphere, hence the name *caisson* disease has been used The "bends" is a popular name for the condition In dysbarism the *percentage* reduction in pressure is a more important factor than the absolute amount of decrease A sudden reduction in pressure to approximately 45% of the pressure with which the subject was equilibrated ordinarily is necessary in order to cause dysbarism A diver who has been exposed to a pressure of, for example, 8 atm usually can be subjected to a quick reduction in pressure to 4 atm without developing symptoms, then after he has been at this pressure long enough to approach equilibrium with the surrounding pressure he can be quickly lowered to 2 atm where he remains for a time and next the pressure can be lowered to 1 atm

The cause of symptoms during rapid decompression is the formation of bubbles of gas in fatty tissues and blood The gas is mainly nitrogen The total amount of nitrogen dissolved in the body when one is exposed to a gas mixture having the same percentage composition as atmospheric air is several times that of O_2 and double that of CO_2. Also, nitrogen diffuses less readily and, unlike O_2, it is not used in metabolic processes When body fluids are saturated with N_2 at a given pressure and the ambient pressure suddenly is lowered, N_2 is evolved in gaseous form just as CO_2 is evolved from carbonated water when the cap is removed from the bottle containing it Bubbles in the blood tend to lodge in the smaller vessels and obstruct the flow of blood Symptoms vary depending upon the sites where the bubbles lodge Fatty tissue usually is the site of bubble formation since N_2 is five times more soluble in oil than in water The involved tissues include subcutaneous tissue, bone marrow, adrenal cortex, myelin sheaths, etc Bubbles in the myelin sheaths of sensory nerves cause pain (bends), paresthesias, itching, etc , and those in myelin sheaths of motor nerves may cause temporary paralysis It has been recognized for many years that obese persons are more susceptible to "the bends" than lean persons

In rapid ascents in airplanes, dysbarism is not a problem up to 20,000 ft since the reduction in pressure is less that 55%, and, in fact, symptoms ordinarily do not appear below 30,000 ft Above this level, however, dysbarism is common The tendency to develop decompression illness is greater the more rapid the ascent The most effective way to prevent or diminish the effects of rapid decompression is to have the aviator breathe 100% oxygen or an oxygen-helium mixture for a time before the flight and thus wash much of the nitrogen out of the body When this is done the aviator should be able to ascend quite rapidly to 40,000 ft (continuing to breathe 100% O_2) without discomfort

The effects of aeroembolism in flyers during rapid ascent take the form most commonly of severe pain in one or more of the large joints and itching of the skin or cutaneous sensations of heat or cold Other more serious symptoms such as paralysis, intense burning pain in the chest, or pulmonary edema may, though rarely, occur. The symptoms are rarely as severe in the decompression which occurs in ascent of altitude, as they are in compressed air illness, since the amount of gas which can be released is not as great

The *explosive decompression* technique which is used to study adjustments to lowered P_{O_2} in the inspired air also is useful in the study of dysbarism A rate of ascent can be simulated in this way which is much faster than any which has been attained even in rockets Many thousands of flying personnel have been subjected to explosive decompression The subject is instructed to breathe with mouth open As the gas in the lungs expands during decompression it rushes out through the open airway Gas in the sinuses also expands suddenly and exhausts through the natural channels, but if these openings are not patent severe pain is produced

Experimental animals have been found to have an astonishingly high tolerance to explosive decompression at simulated high altitudes (from sea level to 50,000 ft —750 to 87 mm Hg) The rate of decompression in the experi-

ments of Whitehorn and his associates ranged from about 1100 mm Hg/sec to as high as 33,000 mm Hg or more/sec That is to say, decompression from 750 mm Hg to 87 mm Hg was brought about in the latter instance in 0 02 sec Then the animal was quickly recompressed and the effects of anoxia thus prevented The chief effects observed in these experiments were distension of the hollow viscera and lungs, fixation of the expanded thorax in the inspiratory position and of the diaphragm in the expiratory position, due to upward pressure of the distended stomach and intestines, sharp but temporary fall in blood pressure with cardiac slowing, hemorrhages into lungs and hollow organs, or rupture of the latter The fall in blood pressure seems to be due to decreased cardiac filling caused by the increased intrathoracic pressure, due in turn to the rise of the diaphragm The cardiac slowing does not occur after bilateral section of the vagus nerves In no instance out of some 700 decompressions on several different species of laboratory animal did a fatality occur from a single decompression The injuries which are most likely to occur from explosive decompression are damage to ears and lungs, though rupture of the stomach or intestine may result

Respiration in a Sealed Cabin

In flight beyond the earth's atmosphere the passengers must be in a sealed compartment The total pressure of gases in the compartment could be reduced quickly to one half that at sea level without causing any difficulty in normal subjects providing the Po_2 were kept up to a level sufficient to prevent arterial hypoxia This could be done if the percentage of O_2 were about 35 to 40 at a cabin pressure of 380 mm of Hg However, it may be argued that it might be preferable to maintain ambient gases at the level normally encountered at sea level The atmospheres within manned aerospace capsules launched by the United States, prior to the fire which resulted in the death of three astronauts, consisted of 100% O_2 at one-third of an atmosphere This results in an oxygen tension of 250 mm Hg in inspired air or well above the 150 mm Hg which normally is inspired at sea level A disadvantage of this may be the greater fire hazard, and the oxygen tension above normal could lead to decreased erythropoiesis The presence of some amount of an inert gas such as nitrogen or helium would help prevent atelectasis and could have other advantages not yet apparent After all, the human body has evolved over thousands of years to survive in a mixture consisting largely of nitrogen At present it is being planned that the occupants of a space station who will spend weeks in an artificial atmosphere will breathe a gas mixture at one atmosphere that has a percentage composition with respect to oxygen and nitrogen not much different from that of the atmosphere The question of the optimal composition of the ambient atmosphere to be maintained in a manned space station was discussed fully in a conference conducted by the Space Science Board (see Carlson, 1967)

Average values for the daily metabolic turnover of a "standard" man are presented in Table 6 6 (compiled by H G Clamann, United States Air Force School of Aviation Medicine, Randolph Field, Texas) Calculations of what supplies must be taken in a sealed cabin for trips of any duration can be made on the basis of such data

Pneumonia

In *lobar pneumonia* the oxygen saturation of the arterial blood varies in different cases from normal to less than 70% The signs and symptoms of anoxia usually appear when the saturation is around 85% Cerebral symptoms, e g , sleeplessness and delirium, cyanosis and dyspnea increase with the oxygen desaturation In lobar pneumonia the carbon dioxide content of the arterial blood is reduced on the average by about 15% (Meakins and Davies) The "alkali reserve," however, is normal or only slightly reduced The blood pH may in some cases be increased From these findings it appears that there exists a partially compensated alkalosis, induced by the hyperventilation (blowing off of carbon dioxide) elicited by reduced oxygen tension of the arterial blood The increased ventilation of the alveoli also increases the percentage of oxygen in the alveolar air, but this cannot, as we shall see presently, increase appreciably the oxygen in the arterial blood.

In *bronchopneumonia* a higher degree of oxygen desaturation of the arterial blood is usually present than in the lobar type The cyanosis may be extreme There is often *retention* of carbon dioxide and a rise in plasma bicarbonate results, to compensate, in part at least, the gaseous acidosis The carbon dioxide

Table 6 6
Daily metabolic turnover
(Man 70 kg RQ = 0 82 Food Protein 80 g , Carbohydrate 270 g , Fat 150 g)

Substance	Input			Output		
	g	%		*g*	%	
Gases	862	24 04	(Oxygen, 603 liters)	982	27 39	(Carbon Dioxide, 496 liters)
Liquid	2200	61 37	(Water)	2542	70 91	(Water)
Solids	500	13 95	(Food)	27	0 75	(Urea)
	23	0 64	(Salts)	23	0 64	(Salts)
				11	0 31	(Uraccounted for)
Total	3585	100 00		3583 + 2830 Kcal	100 00	

content of the arterial blood may be 80 vol % or more

The Causes of Hypoxia in Pneumonia Two main factors concerned in the production of the hypoxia are the passage of blood through unaerated (consolidated) portions of the lung and shallow breathing Typically the blood of pneumonia patients has a normal oxygen capacity and the oxygen dissociation curve at a given carbon dioxide tension is not appreciably different from that of normal persons

In *lobar pneumonia* during the stages of engorgement and red hepatization the alveoli of the affected portion of the lung are poorly aerated Mucus blocks the bronchioles, and the *air spaces* are filled or their walls coated with exudate But a large proportion of the vessels of these unaerated regions are still pervious Consequently, blood traversing such areas must remain poorly oxygenated This blood with a low oxygen saturation and a high carbon dioxide content mixes with blood which has passed through aerated regions The general arterial blood therefore has its oxygen saturation reduced in proportion to the amount of unsaturated blood with which it is mixed (see also "Shunt," p 6-76) When the pneumonic area passes into the stage of gray hepatization the vessels of the affected lobe become obliterated to a large extent and the pulmonary blood then passes through aerated regions (Fig. 6 32). That is, the arterial blood is no longer vitiated by blood from nonaerated areas Therefore in a typical case of lobar pneumonia when the disease is confined to a single large area and the breathing is not shallow, there is little anoxemia at this stage. If however, bronchopneumonic areas coexist, the respiratory functions will be affected as described below (See "Oxygen Therapy," p 6-81)

In *bronchopneumonia*, patches of lung tissue are cut off from their air supply The fine bron



Fig 6 32 X-ray photograph of lung (injected with barium) from a case of lobar pneumonia (after Gross) *Upper right* Consolidated area—red hepatization The main vessels are constricted and the finer vascular structure is less dense than in the normal lung *Lower right* Consolidated area—gray hepatization The main vessels are patent but the finer vessels have been occluded *Lower left* Healthy portion of lung except for compensatory congestion, the vessels are dilated

chioles become plugged with mucus, groups of alveoli become filled with exudate, and the alveolar walls are edematous and thickened Yet, obliteration of the vessels to any great extent does not occur Blood continues to flow through unaerated areas This blood which has a low Po_2 on mixing with that from aerated alveoli lowers the saturation of the general arterial blood

Shallow Breathing In pneumonia breathing is frequently very rapid and shallow Instead of the tidal air being around 500 ml as in health it may be reduced to 250 ml or less It will be recalled that about 150 ml are required to fill the anatomical dead space, therefore only 100 ml will enter the air sacs of the healthy parts of the lung The expansion of the lungs is not equal in all its parts The alveoli toward the hub of the radiating rays expand less than those near the periphery Those parts, such as the apex, which are indirectly expanded have even in health a tendency to be ventilated less than those which are directly expanded In shallow breathing, these differences are greatly exaggerated Although the tidal volume is only half the normal, the total quantity of air breathed per minute (minute volume) is, as a result of the increased respiratory rate, much greater than normal Since a proportion of the alveoli are very poorly ventilated or not at all, owing to the unequal expansion of the lung, those in other areas tend to be overventilated The O_2 tension in the latter is therefore raised But so far as the oxygenation of the blood is concerned, the overventilation of some alveoli cannot make up for the underventilation of others We know that the hemoglobin is nearly saturated already at the ordinary alveolar oxygen tension of 100 mm Hg As already mentioned the dissociation curve of hemoglobin in pneumonia does not differ appreciably from that in health, and the most, therefore, that could be expected from a rise in the alveolar O_2 tension would be an increase of 2% or less in oxygen saturation of the blood traversing overventilated regions and a slight increase in the amount of O_2 held in simple solution In other words, the blood flowing through the poorly ventilated parts of the lung will have a low saturation since the O_2 tension is low, while that flowing through the overventilated parts will be little above the normal The net result will be a low oxygen saturation of the mixed arterial blood

Matters are different in the case of CO_2 elimination The shape of the CO_2 dissociation curve which shows a progressive slope throughout the entire range of CO_2 tensions is quite unlike that for oxyhemoglobin (Fig 6 33) The greater total ventilation results in a lowering of CO_2 tension in the overventilated parts of the lung and, consequently, in a "blowing off" of CO_2 from the blood circulating through these regions CO_2 is retained in the blood circulating through poorly ventilated regions In the patient with lobar pneumonia the amount of CO_2 blown off may exceed that retained, the net result will be a lowering of the CO_2 content of the arterial blood In bronchopneumonia a larger proportion of the pulmonary blood circulates through nonaerated areas As a consequence, CO_2 retention is greater, and a normal or a higher than normal CO_2 content of the blood is more usual

The Cause of Shallow Breathing in Pneumonia Pleuritic pain, by restricting the respiratory excursions, may result in this type of breathing In other instances it may be of a reflex nature resulting from the inflammatory process which, through a reduction in the distensibility of the pulmonary tissue perhaps exalts the sensitivity of the afferent vagal endings

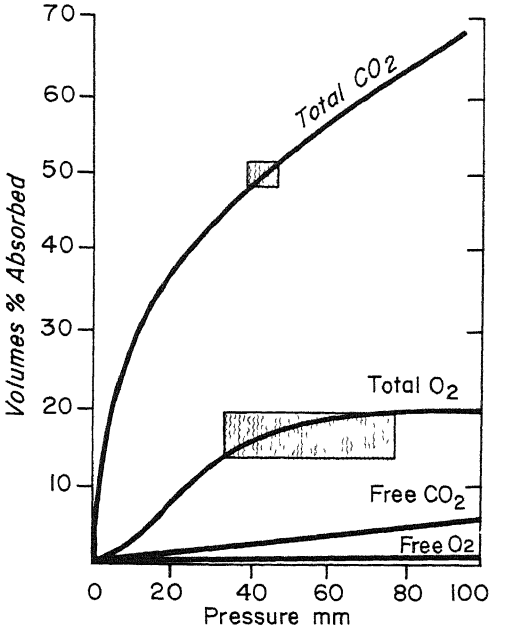


Fig 6 33 Carbon dioxide and oxygen dissociation curves The small rectangles indicate the extent of the variation of the O_2 and CO_2 of the subject s blood when at rest (From L J Henderson 1928)

in the alveolar walls Thus the inspiratory movement possibly is terminated before a full excursion has been completed (see Hering-Breuer reflex, p 6-47) In support of a reflex origin may be mentioned the experiments of Dunn and of Binger, Brow, and Branch who produced this type of breathing in animals by the intravenous injection of potato starch granules These, acting as small emboli, plugged the pulmonary capillaries The rapid shallow breathing was immediately abolished by section of the vagi or prevented if the nerves had been cut before the injection Also, in pneumonia it has been found that oxygen inhalations, even though they may restore the oxygen saturation of the arterial blood to normal, do not necessarily abolish the shallow breathing This constitutes further evidence for the existence of a nervous element in the production of this type of breathing Shallow breathing may also result from other diseases involving the alveoli, e g , inflammation by irritant gases (phosgene and chlorine), pulmonary edema, miliary tuberculosis, and pulmonary emboli, which would be expected to stimulate afferent nerve endings

Bronchial Asthma

This is a paroxysmal disease in which acute oxygen want is caused by a spasm of the smooth muscles of the finer bronchioles Edema of the bronchiolar mucosa is probably also present The alveoli are poorly ventilated, and some may be completely cut off from their air supply The high percentage of carbon dioxide and low percentage of oxygen in the alveolar air result in a low oxygen saturation of the arterial blood and the retention of carbon dioxide The acidosis related to CO_2 retention is met by the excretion of a highly acid urine and a rise in the "alkali reserve " An intense plum-colored cyanosis may develop The continued stimulating effect of oxygen want and carbon dioxide excess upon the respiratory mechanisms causes severe dyspnea

Difficulty is experienced both in inspiration and expiration, but since there is a natural tendency for the bronchioles to narrow during expiration and dilate during inspiration, the greatest respiratory effort is exerted during expiration The respiratory muscles contract with great force, and the accessory muscles of respiration are brought into play The expiratory muscles compress the chest, the abdominal muscles contract in the attempt to squeeze the air from the lungs The intrapulmonary pressure is greatly elevated, and the air escapes through the constricted tubes with a distinct wheezing sound Due to the difficulty and prolongation of the expiratory phase, normal deflation of the lungs cannot occur before the next inspiration ensues The lungs, therefore, remain somewhat expanded even at the end of expiration That is, during the asthmatic paroxysm a very large volume of residual air is present in the lungs (Fig 6 34) The tidal volume and the vital capacity are greatly reduced The changes in volume of the overdistended lung are small and not commensurate with the excursions of the thoracic walls As a consequence, the high value of the intrathoracic negative pressure induced during inspiration causes the structures at the root of the neck to be drawn toward the thoracic cavity to take up the space which the lungs are unable to fill During expiration the veins of the neck and face become engorged The restricted movements of the lung also greatly reduce the effect of mechanical mixing upon the lung air, the process of diffusion being depended upon to a larger extent for the freshening of the alveolar air

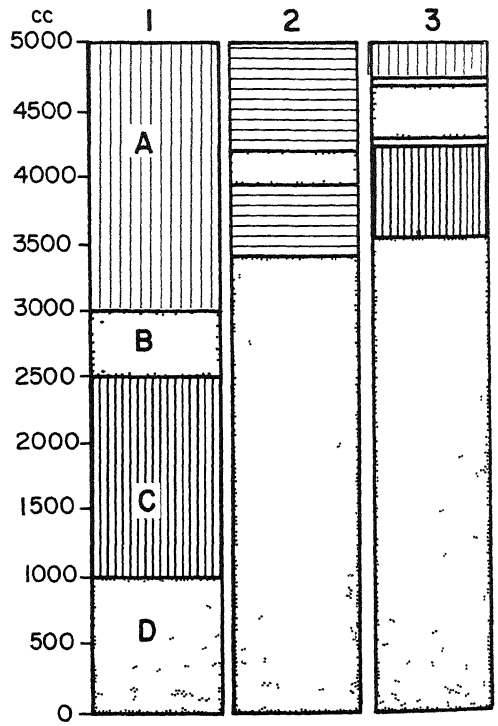


Fig 6 34 Diagram showing subdivisions of the lung air in asthma (2) and emphysema (3) compared with the normal (1) *A* Inspiratory reserve volume *B* tidal volume *C* expiratory reserve volume and *D* residual volume

CAUSATION

Asthma often shows a strong hereditary tendency The bronchiolar spasm may be of a *reflex nature* and due to the stimulation of hypersensitive afferent vagal endings in the larynx, or of trigeminal fibers by some nasal abnormality, the bronchoconstrictor impulses travel via efferent vagal fibers On the other hand, the bronchiolar spasm may be an *allergic phenomenon*, i e , the result of sensitization to some foreign protein This is the most common cause of the condition It is then frequently associated with other allergic conditions, e g , hay fever, urticaria, or eczema either in the patient himself or in members of his family The foreign protein may be inhaled Pollens of various grasses and flowers, the dandruff of animals, e g , horse, cat or dog, or feathers are among the most common excitants, or the exciting cause may be some kind of food or the protein of bacteria within the respiratory tract itself The sensitivity of certain individuals to foreign proteins presents many features resembling those of anaphylactic shock in animals Sudden death may result from the injection of horse serum (e g , diphtheria antitoxin or antitetanic serum) into an asthmatic subject A guinea pig when injected with a protein to which it has previously been sensitized dies rapidly from anaphylactic shock The bronchiolar muscle is strongly contracted The air is trapped so that the lungs are maximally distended and do not collapse when the thorax is opened Even when the pulmonary tissue is deeply incised the air does not escape from the distended lung This manifestation of anaphylaxis is associated with (and perhaps due to) the liberation of histamine It is well known that anaphylaxis and histamine administration produce almost identical effects in the guinea pig (Fig 6 35)

The *treatment* of asthma resolves itself into the relief of the paroxysm and the removal of the underlying cause Epinephrine or ephedrine acts by inhibiting the bronchiolar muscle during the attack Atropine blocks the action of the bronchoconstrictor (vagal) fibers In the allergic form of the disease every effort is made to identify the offending protein Antihistaminic drugs have been employed with varying degrees of success

Chronic Emphysema

The lungs in emphysema (Greek *em* + *physema*, a blowing) are in a state of extreme distention as a result of the enlargement of the air sacs The latter, however, show fewer alveoli in their walls owing to the atrophy of the interalveolar septa Contiguous air sacs within a lobule coalesce or even adjacent lobules may fuse to form large air spaces For this reason the total respiratory surface is reduced The alveolar and capillary walls become thickened and the interstitial pulmonary tissue increased Many capillaries become occluded The pulmonary elastic tissue is reduced in amount so that the lungs, when removed from the thorax, do not collapse normally but remain in an overexpanded state The peripheral lobules, which in health expand to the greatest extent, are those mainly affected in emphysema, the enlarged lobules appearing as blebs upon the surface of the lung

The chest is what is known as barrel-shaped The ribs are more horizontal than normally, the thoracic spine is bowed backwards (kyphosis) so that the anteroposterior diameter of the chest is as great as or exceeds the transverse The position of the chest is one of nearly full inspiration The midposition of the diaphragm is at a much lower level than usual and its excursions above and below this level are restricted The respiration is therefore mainly costal In some instances the diaphragm is practically fixed or indeed may be drawn up during inspiration (paradoxical movement) The residual air is

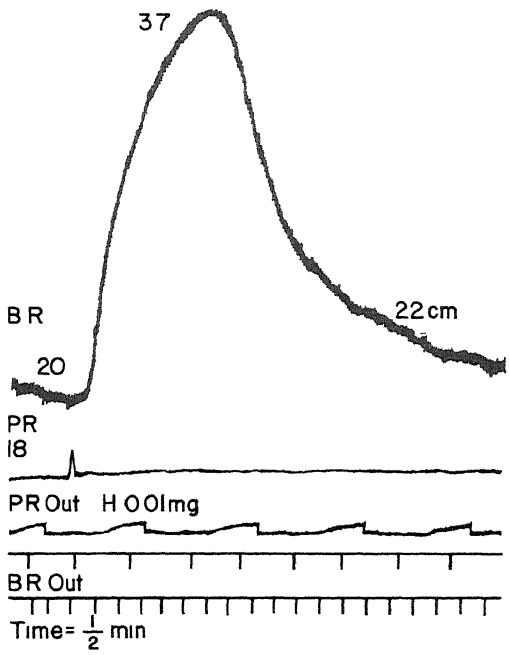


Fig 6 35 Guinea pig Reaction of bronchial muscle to 0 01 mg histamine

two or three times the normal, and the inspiratory reserve volume is reduced The tidal volume is normal or only moderately reduced and the vital capacity (Fig 6 34) is lowered by from 20 to 60% Owing to the loss of the elasticity of the lung, expiration is no longer simply a passive movement but is aided by a forcible contraction of the expiratory muscles When the patient is asked to make a deep inspiration, he does not expel all the air during the next expiration Several respirations occur before the chest returns to its original size Expiratory reserve volume is greatly decreased, and residual volume is markedly increased (Fig 6 34) These phenomena are due to the fact that the inelastic lungs have been overstretched and are brought back to their original volume with difficulty

In emphysema there are arterial hypoxia and retention of carbon dioxide The oxygen saturation of the arterial blood runs from a little below normal to around 75% or even less The carbon dioxide tension in the alveolar air in well-marked cases is from 50 to over 60 mm Hg (7 to 8%) and the carbon dioxide content of the arterial blood is correspondingly high

Owing to the obliteration of many capillaries in the lungs, the resistance in the pulmonary circuit is increased, thus a greater burden is put upon the right ventricle

The cause of the impaired gaseous exchange is not altogether clear Thickening of the alveolar and capillary walls and the obliteration of capillaries have been considered to be factors Yet if this were so one would not expect the retention of carbon dioxide which, owing to its greater solubility (30 times that of oxygen) has a much higher rate of diffusion through the pulmonary membrane, to be so much more pronounced than the anoxia The sharp rebound at the end of inspiration which occurs in the healthy lung causes mechanical mixing of the lung air and is an important factor in the efficient ventilation of the alveoli The absence of this effect in the emphysematous lung and its greater dependence, in consequence, upon the slower process of diffusion is probably an important factor leading to the defective aeration of the blood According to Christie, the impaired gaseous exchange is due mainly to the fact that, as a result of the loss of elasticity, the intrathoracic pressure is not distributed evenly throughout the lung As a consequence, the outlying alveoli which are largely functionless with obliterated vessels are ventilated to a greater extent than the relatively healthy ones more centrally placed Owing to the shapes of the respective dissociation curves such underventilation of the functioning alveoli will tend to have a greater effect in preventing the elimination of CO_2 than in interfering with the absorption of oxygen Other factors which may be concerned, are the slower rate of diffusion of CO_2, owing to its larger molecule, in the alveolar air, and the increased residual volume

The red cell count, hemoglobin content, and consequently the oxygen capacity of the blood, are increased above the normal in emphysema The cyanosis (p 6-78) is often pronounced, yet the patient's dyspnea is less than might be expected from his color and from the carbon dioxide retention which exists This is explained by the well-established fact that in emphysema the respiratory center is less sensitive to carbon dioxide A normal person when breathing a carbon dioxide rich mixture (8%) increases his pulmonary ventilation by 300% or more, the breathing of the emphysematous patient, on the other hand, shows relatively little change as a result of breathing a much stronger mixture (see Fig 6 36)

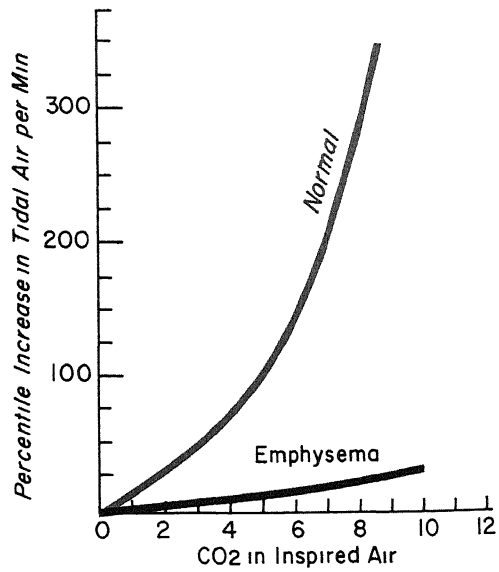


Fig 6 36 Chart showing the percentile increase in tidal air per minute as the percentage of inspired carbon dioxide is raised Note that when the normal subject inspires air containing 8% carbon dioxide the tidal volume is increased about 300% whereas in the emphysematous subject it is increased only about 25% (After R W Scott 1920)

CAUSATION

Two factors are concerned in the production of emphysema *a)* reduction in the elastic tissue of the lung, and *b)* increased distention of the alveolar spaces

It is very questionable whether, in the absence of some abnormality of the lung tissue itself, emphysema can result from increased intrapulmonary pressure, such as occurs in those following certain occupations, e g , glass blowers and the players of wind instruments The study of groups of men following such occupations does not indicate that emphysema is produced in this way Emphysema has, however, been induced in animals by stenosis of the trachea or bronchi or by the insertion of a valved apparatus into the trachea which allowed the free ingress of air but offered an obstruction to expiration

Asthma and chronic bronchitis frequently are forerunners of emphysema In the former condition the spasm of the bronchiolar muscle exerts a valvelike action In chronic bronchitis mucous plugs would act similarly Coughing, it has been supposed, places a strain upon the alveolar walls But during the phase of coughing when the glottis is closed the alveolar walls are supported When the glottis opens, the air escapes from the alveoli if the obstructing material has been dislodged, and no strain upon the alveolar wall would result If, however, the air remains entrapped its sudden re-expansion (rebound), as the pressure in the surrounding pulmonary tissue falls at the end of a cough, may injure the alveolar membrane and start the emphysematous process

What may be termed a physiological or *compensatory* emphysema occurs when part of the pulmonary surface is reduced as by the collapse of a part or the whole of one lung This is more in the nature of an hypertrophy A similar enlargement of both lungs occurs at high altitudes

A type of emphysema also occurs in the elderly—*senile or postural emphysema*—and is secondary to the change in the shape of the thorax which becomes more barrel-shaped and increased in capacity The lungs enlarge to fill the increased space This condition is associated with few symptoms There is little reduction in the vital capacity, and the oxygen saturation of the arterial blood is practically normal

Atelectasis

Any condition which lowers the pressure within the alveoli or increases the pressure upon the lung surface, may lead to collapse of the lung Thus pleural effusions, pneumothorax, tumors, etc , pressing from without or the isolation of the alveoli from their air supply by the obstruction of a bronchus will result in collapse of the lung or of the portion of the lung affected Atelectasis (Greek *ateles*, incomplete, *ektasis*, distention) is also the term applied to the condition in the newborn in which, as a result of the blockage of a bronchus or of a group of bronchioles by mucous secretion, or owing to weak inspiratory movements, a portion of the lung fails to become distended with air

Absorption of Air from Isolated Alevoli and Other Closed Spaces When a bronchus or bronchiole in a previously distended lung is obstructed the imprisoned air soon becomes absorbed from the affected alveoli Collapse of the air sacs cannot take place until this has occurred Absorption is brought about in the following way as pointed out by Henderson The air in the isolated alveoli has a total pressure of 760 mm Hg The partial pressures are in round numbers, O_2, 100 mm , N_2, 570 mm , CO_2, 40 mm , and aqueous vapor, 47 mm In the venous blood the total pressure is 703 mm , the nitrogen and aqueous vapor being the same as that of the alveolar air, but the partial pressure of oxygen is only 40 mm and of carbon dioxide 46 mm An interchange of the latter gases therefore occurs between the alveolar air and the venous blood It might be thought that the imprisoned air would then be in equilibrium with the blood and no further absorption occur But the alveolar air loses more oxygen than it gains carbon dioxide whereupon its volume is reduced The atmosphere acting upon the body surface and through the yielding soft tissues compresses the air so as to maintain its total pressure practically constant at 760 mm As a result of the absorption of oxygen the percentage and consequently the partial pressure of carbon dioxide and of nitrogen are increased These gases now diffuse into the blood The volume of the alveolar air is further reduced thereby, but its total pressure still remains unaltered The percentage and therefore the partial pressure of oxygen rises and more of this gas passes into the blood The process continues in this manner until no air remains, and the walls of the original space are ultimately approximated by the pressure of the atmosphere Air is absorbed from the pleural cavity or from any other closed cavity of the body in the same way (Henderson and Henderson, 1932) The absorption of the air confined within the pleural cavity permits the lung, if the air pressure had caused its collapse, to re-expand

The collapse and shrinkage of the lung which results from blockage of a bronchus increases

the intrathoracic negative pressure, since the closed thoracic box is less completely filled The diaphragm is therefore drawn upwards and uncollapsed portions of the lungs are expanded to a greater extent (compensatory emphysema) in order to fill the unoccupied space

Shunt

In the fetus the vessels of the airless lungs are bypassed by the greater part of the blood brought to the heart That is to say, most of the blood is shunted to the arterial side through channels which normally close at, or shortly after, birth It is clear that if one or more of these channels should persist after birth, or if the interventricular septum is defective so that a large part of the venous blood does not traverse the lungs, but is short circuited from the right to the left heart or directly into the aorta, the oxygen saturation of the arterial blood will be seriously reduced Arterial hypoxia evidenced by cyanosis (p 6-78) and dyspnea (p 6-59) will result (Fig 6 37) These will be especially pronounced upon exertion since the unsaturation of the polluting venous blood will be thereby increased

The Fetal Circulation Before the congenital abnormalities are described, the fetal circulation and the circulatory readjustments which occur at birth, or shortly thereafter, will be briefly reviewed

In the fetus the oxygenated blood from the placenta is carried by the *umbilical vein* to the liver, where it separates into two streams, one of which is distributed to the left two-thirds or so of the liver, the remainder passes directly via the *ductus venosus* into the inferior vena cava, thus mixing with blood returning from the lower part of the body The portal vein supplies the rest of the liver That portion of the blood from the umbilical vein which has traversed the vessels of the liver is also delivered by the hepatic veins into the inferior vena cava The hepatic veins are a number of small vessels and two much larger ones, the latter form a junction with the ductus venosus just before the latter joins the inferior vena cava The blood in the vena cava beyond this point is, therefore, partly oxygenated (from placenta via umbilical vein and ductus venosus) and partly reduced, having traversed the tissues of the lower limbs, intestines, and liver The caval blood, upon reaching the right auricle, separates into two unequal parts The much larger stream passes directly through the foramen ovale (*via sinistra*) into the left auricle, the smaller stream enters the right ventricle The blood which has reached the left auricle, after being joined by a small but by no means inconsiderable volume of blood returned from the pulmonary tissue, flows into the left ventricle, and is discharged into the aorta and distributed to the brain by the brachiocephalic artery, and to the myocardium by the coronary arteries The remainder flows down the aorta

The blood returned from the head and entering the right auricle by the superior vena cava, together with the small fraction of blood of the inferior vena cava which was not directed through the foramen ovale, enters the right ventricle and is ejected into the pulmonary artery A smaller part, yet, as shown by Franklin and his associates, more than has been generally supposed is conveyed to the lungs by two branches of the latter vessels The main stream, discharged from the right ventricle, passes by a wide vessel, the *ductus arteriosus*, into the aorta, and mixing with the blood coming from the left ventricle, is distributed throughout the body That portion which is supplied to the lower limbs is returned by two vessels—the *umbilical arteries*—to the placenta

During birth, or very shortly after, the three umbilical vessels close (even though the cord has not been tied or broken) by the contraction of smooth muscle in the vascular walls The flow through the ductus venosus ceases as a result of contraction of a smooth muscle sphincter at the point where it leaves the umbilical vein This occurs in the sheep fetus, according to Franklin and his associates, in from 5 to 25 min after the umbilical vessels have closed Within 5 min or so after the commencement of breathing the foramen ovale closes, and a little later (apparently never before the obliteration of the foramen ovale) the ductus arteriosus becomes occluded All the blood from the right side of the heart is nor-

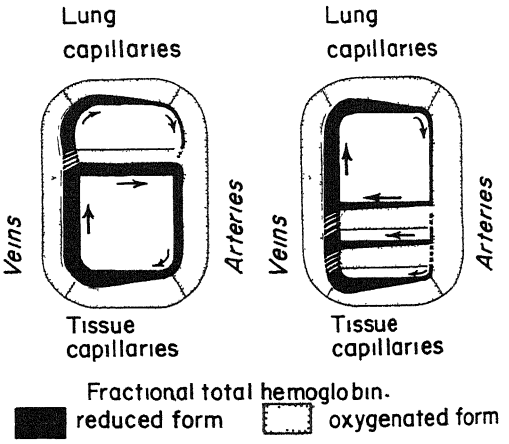


Fig 6 37 Diagram (*left*) showing the proportion of oxyhemoglobin to reduced hemoglobin in different parts of the circulation in an instance where a portion of the blood passes through unaerated channels (shunt) from the venous to the arterial system Righthand diagram represents a case in which the oxygen unsaturation of the blood is abnormally high in a part of the peripheral capillaries but normal in the arterial blood (hypokinetic type of anoxia) (After Lundsgaard and Van Slyke 1923)

mally from now on directed through the lungs Franklin and his associates found, however, in the case of the newborn lamb, that should the general condition of the later seriously deteriorate, the ductus arteriosus may open again The mechanism of closure of the ductus appears to be initiated in some way by the action of the oxygenated blood upon the vascular muscle and not through a nervous mechanism In the lamb this channel has never been observed to close before occlusion of the umbilical vessels (Barclay et al , 1945)

CONGENITAL CARDIAC DEFECTS

Patent ductus arteriosus, or foramen ovale, interventricular septal defects, stenosis of the pulmonary artery and coarctation of the aorta (narrowing of aorta in the neighborhood of the ductus arteriosus) are among the congenital defects most commonly met

The development of the direct Fick method (sect 3, chap 7) (utilizing catheterization of the right side of the heart) for determining the cardiac output has provided means of the greatest value in diagnosing congenital defects of the circulation It is now possible to obtain samples of blood for gas analysis from the right ventricle, the pulmonary artery, and many sites in the systemic circuit, including the left ventricle (See Fig 6 38)

Patent Foramen Ovale or Other Defect in the Interauricular Septum In patency of the foramen ovale, the opening is guarded by a valvelike membrane which prevents blood from passing from the left to the right auricle, and none passes in the opposite direction unless the pressure in the right auricle is high With other defects of the interauricular septum, a shunt from right to left with cyanosis may occur, but if there is mitral stenosis and a large defect (*Lutembacher's syndrome*) the high pressure in the left auricle causes a shunt from left to right

Patent Ductus Arteriosus Blood enters the pulmonary system from the arterial side, the oxygen content in the pulmonary artery is increased, but is normal in the right auricle and ventricle

If a *defect of the interventricular septum* exists alone, the oxygen content of the blood in the right ventricle and pulmonary artery is increased but not of that in the right auricle

Neither of these last two defects, if it exists alone, will cause cyanosis, since the shunt is from left to right (arteriovenous shunt) Rarely, however, in the case of a patent ductus arteriosus, the direction of flow is reversed temporarily at certain times (during crying or coughing), blood passing from the pulmonary artery into the aorta

Often the abnormalities are multiple For example, a patent foramen ovale, together with stenosis of the pulmonary artery, results in the passage of venous blood to the arterial side, i e , through the auricular defect, slight or moderate cyanosis is seen The combination of congenital defects, originally described by John Hunter and later by Fallot, and now generally known as the *tetralogy of Fallot*, consists of, *a*) an interventricular septal defect, *b*) narrowing of the pulmonary artery, or stenosis of its orifice, and *c*) dextroposition of the aorta, the orifice of this vessel being shifted to the right to straddle the septal defect *d*) The fourth element making up the tetrad is right ventricular hypertrophy The blood flow through the lungs and the pulmonary pressure are reduced Owing to the high pressure developed in the right ventricle blood is driven through the defect into the overriding aorta If the pulmonary stenosis is very severe, the ductus arteriosus must be patent in order to provide sufficient blood flow through the lungs to support life Cyanosis and dyspnea are usually pronounced Blalock and Taussig have devised an operation for the correction of the chief effects of these conditions It consists of anastomosing a large branch of the aortic arch (usually carotid or subclavian) to the right or left pulmonary artery

Another combination of defects has three features of the tetralogy of Fallot, namely, defect of the interventricular septum, dextroposition of the aorta and right ventricular hypertrophy, but pulmonary stenosis is absent It is known as *Eisenmenger's complex* The shunt is from left to right in early life, but later, when as a result of the large volume of blood entering the right ventricle from the left side as well as through the normal channels this chamber enlarges and hypertrophies, the high pressure developed may force venous blood into the overriding aorta, cyanosis then appears

In those congenital cardiac anomalies showing

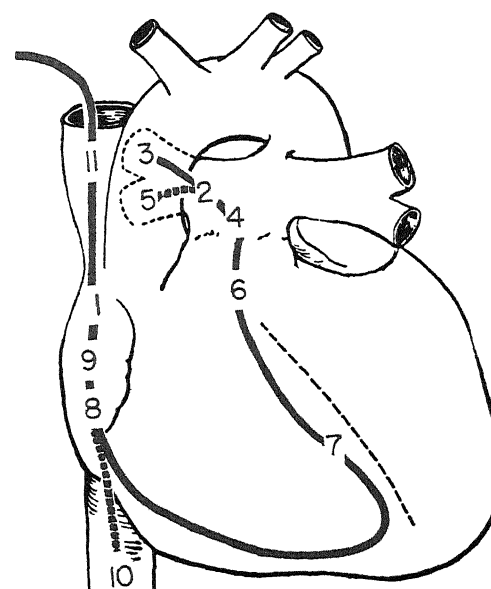


Fig 6.38 Numbers from 1 to 11 indicate the positions from which blood samples may be taken by means of the intracardiac catheter (After Groom et al 1948)

cyanosis, the arterial oxygen saturation is found to be, during rest, from 45% in severer degrees of cyanosis to 90% in the milder cases The red cell count, however, is likely to be increased, and the hemoglobin concentration may be high as 26 g % The arterial oxygen content, i e , the volumes of oxygen/100 ml of blood, in those with the higher hemoglobin concentration is, consequently, much higher than normal and may reach a value of 29 vol %, as against the normal maximum of 22 vol % (see "Cyanosis," below) Exercise, which in normal subjects causes little or no change in arterial oxygen saturation, produces a profound fall in the congenital heart cases, amounting in some to a reduction of over 30%

Carbon Monoxide Poisoning

References to carbon monoxide poisoning are contained in the earliest medical writings This gas was used by the Greeks and Romans for the execution of criminals and as a means of committing suicide Today, carbon monoxide is the most important gaseous poison against which physicians have to contend

The affinity of hemoglobin for carbon monoxide is approximately 300 times its affinity for oxygen Therefore, when the atmosphere contains only a very small percentage of CO, the hemoglobin takes up the poisonous gas and in so doing loses its ability to combine with an equivalent volume of oxygen

Recovery from carbon monoxide poisoning is usually complete when the exposure has not been too long or the concentration too high It is important that the carbon monoxide hemoglobin should be broken up as soon as possible since injury to the tissues is produced by the anoxia It is possible to displace the carbon monoxide by oxygen if the tension of oxygen is sufficiently high and that of carbon monoxide low Mixtures with a high percentage of oxygen and from 6 to 7% carbon dioxide combined with artificial respiration are used in the treatment of CO poisoning The carbon dioxide, as well as acting as a respiratory stimulant, reduces the affinity of hemoglobin for carbon monoxide Recently, administration of oxygen under pressures greater than 1 atm has been advocated by Smith (1965) Carbon monoxide, as well as its action in displacing oxygen from hemoglobin, has histotoxic properties, inhibiting the tissue respiratory enzymes This fact has no practical bearing, however, upon carbon monoxide poisoning in man, for the concentrations at which such action occurs is many times greater than could ever occur in the body

Cyanide Poisoning

No interference with oxygen carriage is caused by cyanide, but there is inhibition of tissue respiration Cyanide inhibits the action of cytochrome oxidase carbonic anhydrase and probably of other enzyme systems It follows from this that any antidote for cyanide poisoning must have one of two actions Either the cyanide must be removed or detoxified or the inactivated catalyst must be replaced Methylene blue does act as a catalyst for certain biological oxidations, and this led to its use as an antidote for cyanide poisoning When methylene blue is given to the intact animal, there is a marked rise in body temperature which is due to increased metabolism Sahlin in 1926 provided the first experimental demonstration that methylene blue antagonizes the action of cyanide in the intact animal He used rats and the observation has been confirmed on other animals—dogs, rabbits, mice The evidence suggests that methylene blue acts by removing the cyanide from tissue Methylene blue and cyanide do not combine directly, but methylene blue forms methemoglobin which combines with the cyanide to form cyanmethemoglobin The cyanmethemoglobin is relatively nontoxic and is broken down slowly, the detoxification probably being brought about by conversion of the cyanide to thiocyanate (Smith and Malcolm) Other substances which form methemoglobin (Hug, 1933, Wendel, 1933) such as sodium nitrite, amyl nitrite, pyrogallol, etc , are also effective in the treatment of cyanide poisoning Such methods of course are limited by the quantity of hemoglobin that can safely be converted to methemoglobin

A number of sulphur compounds have been found effective in cyanide poisoning Chen, Rose, and Clowes (1934) showed that sodium thiosulphate and sodium tetrathionate may protect dogs against as much as three times the lethal dose of cyanide

Cyanosis

Cyanosis (Greek, *cyanos*, blue) may be defined as the diffuse, dusky, or bluish color of the skin and mucous membranes caused by the presence in the blood of the superficial capillaries (subpapillary venous plexus) of reduced hemoglobin above a certain definite amount The presence of abnormal compounds of hemoglobin, e g , methemoglobin and sulfohemoglobin, resulting from the action of various toxic substances, causes a type of cyanosis (enterogenous cyanosis), but these will not be considered here (see sect 4, chap 4) Cyanosis may result from anoxia of either the arterial or hypokinetic type It obviously cannot occur in the anemic type, which is due essentially to a low hemoglobin concentration, in the histotoxic type in which the hemoglobin gives up less of

its oxygen store than in health, nor in the arterial and hypokinetic types if a severe grade of anemia exists The retention of carbon dioxide in the blood has no *direct* effect upon the production of cyanosis The blue color of the skin depends fundamentally upon the *absolute* amount of reduced hemoglobin in the capillary blood and *not* upon the *relative proportions* of reduced hemoglobin and oxyhemoglobin For example, in anemia the hemoglobin content of the blood may be only 20% of the normal In the capillary blood all of this could be in the reduced form, yet cyanosis would not result, since the absolute amount of reduced hemoglobin (i e , "blue" pigment) would be insufficient to produce any blue discoloration On the other hand, in polycythemia the hemoglobin may be 100% above normal Cyanosis will occur when the hemoglobin of the capillary blood is only 20% reduced, for the absolute concentration of reduced hemoglobin will then be raised to threshold value The greater quantity of the bright-colored oxyhemoglobin present exerts little or no influence, that is, it does not, as might be expected, tend to neutralize the color effect of the reduced hemoglobin

Normal blood contains about 15 g of hemoglobin/100 ml Lundsgaard found that the capillary blood must contain approximately 5 g of reduced hemoglobin/100 ml before cyanosis will appear When fully saturated (20 volumes %), ¾ g of hemoglobin will take up 1 ml of oxygen Five grams of hemoglobin, therefore, hold about 6 7 ml of oxygen, and 5 g of reduced hemoglobin are formed when the blood contains 13 3 volumes % of oxygen Cyanosis may therefore be expected to appear when the blood in the capillaries is on the average around 7 volumes % unsaturated As a result of certain modifying factors the precise level of capillary unsaturation at which cyanosis makes its appearance varies in different cases between 6 and 7 volumes per cent of oxygen

The oxygen unsaturation of the capillary blood does not of course occur abruptly at the arterial end but is progressive from point to point along the course of the vessel The loss of oxygen may be uniform from the arterial to the venous end of the capillary as shown in curve 1, Figure 6 39, or the desaturation may occur mainly toward the venous end (curve 2) when the capillary blood would approximate arterial blood in its content of reduced hemoglobin Under other circumstances the greatest oxygen loss may occur toward the arterial end (curve 3) when the unsaturation of the capillary blood throughout would approach that of venous blood It is difficult to obtain data from which the true curve may be drawn The simplest of these curves (curve 1) is assumed and the average unsaturation of the capillary blood is taken as midway between that of arterial and venous bloods respectively Thus

$$\tfrac{1}{2}(A + V) = C$$

where A and V represent the unsaturation in volumes per cent of arterial and venous bloods, respectively and C the average unsaturation of the capillary blood

For example, in a normal person the arterial unsaturation is 0 5 volume % (19 5 volumes % saturation), another 5 vol % are given up in the capillary, the venous unsaturation is therefore 5 5 volumes % So

$$\tfrac{1}{2}(0\ 5 + 5\ 5) = 3\ 0 \text{ volumes \% average unsaturation of capillary blood}$$

$$(3\ 0 \times 0\ 75) = 2\ 2 \text{ g reduced Hb}$$

We have seen that the average unsaturation of the capillary blcod must be between 6 and 7 volumes % (approximately 5 g of reduced Hb) before cyanosis appears This degree of unsaturation of the capillary blood may be brought about either by an increase in the arterial unsaturation (arterial type of anoxia) or as a result of a greater amount of oxygen being lost from the blood in its passage through the capillaries (hypokinetic type of anoxia) That is, by an increase in the venous unsaturation (V) alone

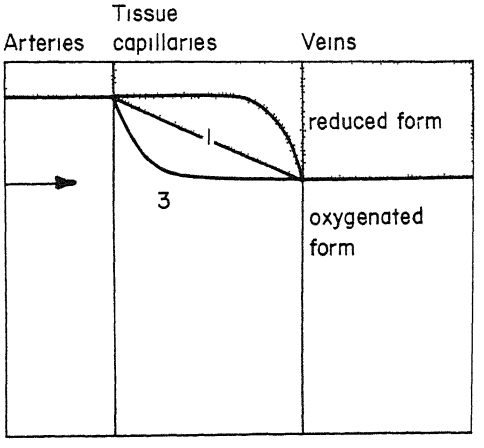


Fig 6 39 Diagram showing hypothetical variations in the shape of the reduction curve of oxyhemoglobin during the passage of the blood through tissue capillaries (After Lundsgaard and Van Slyke 1923)

In order to produce an average unsaturation of from 6 to 7 volumes % in the capillary blood the arterial unsaturation would need to be from 4 to 4 5 volumes %, or the unsaturation of the venous blood—that in the arteries being normal —11 to 13 volumes %. Meakins and Davies (1920) found that when the veins of the arm were obstructed in normal individuals cyanosis was just detectable when the venous blood reached an unsaturation of 11 4 volumes %.

These facts may be clarified by examples. If the blood as it leaves the lungs contains only 15 5 volumes % of oxygen, i e , if it has an unsaturation of 4 5 volumes % and the tissues abstract the usual quantity of oxygen, namely 5 volumes %, the unsaturation of the venous blood will be 9 5 volumes % and that of the capillary blood

$$\tfrac{1}{2}(4\ 5 + 9\ 5) = 7 \text{ volumes } \%$$

$$(7 \times 0\ 75 = 5\ 2 \text{ g Hb})$$

If on the other hand the arterial unsaturation is around the normal value of 0 5 volume % but as a result of slowing of the circulation each unit volume of the blood gives up a greater quantity of oxygen, and the venous unsaturation is increased to say 13 volumes %, the average unsaturation of the capillary blood will be

$$\tfrac{1}{2}(0\ 5 + 13) = 6\ 7 \text{ volumes } \%$$

In either of these instances slight cyanosis would be expected to appear.

Let us now consider what occurs when the hemoglobin percentage is above or below the normal value.

In *anemia* the amount of hemoglobin is below normal, and the oxygen capacity of the blood is correspondingly lowered. If the hemoglobin content be only 30% of the normal, the volumes per cent of oxygen in the arterial blood, though the latter be fully saturated, will be only about 6 volumes %, i e , the quantity of hemoglobin in 100 ml of blood is less than 5 g. It is obvious that such a person could not become cyanotic even if all his hemoglobin were in the reduced state. A patient whose hemoglobin was, say, 60% (oxygen capacity 12 volumes %) would, like one with a normal hemoglobin content, become cyanotic when his arterial blood reached an unsaturation of about 4 5 volumes %, that is when his capillary blood had an average unsaturation of around 7 volumes %. But the oxygen want of the anemic subject would be greater than that of the subject with a normal hemoglobin content since in the case of the former 4 5 volumes % constitute nearly 40% of the oxygen capacity of his blood. In a less enlightened age when bleeding was resorted to for the relief of cyanosis, the brilliant success of that procedure is not to be wondered at—nor that the patient died. In the hypokinetic type of anoxia the unsaturation of the venous blood of a subject with 60% hemoglobin content could not reach the value necessary to produce cyanosis until all the hemoglobin was in the reduced state. Thus ½(1 + 12) = 6 5 volumes % unsaturation of the capillary blood.

In *polycythemia*, in which the hemoglobin content is, say, double the normal (oxygen capacity 40 volumes %), cyanosis occurs in very mild degrees of anoxemia. At an arterial unsaturation of 4 5 volumes % the polycythemia subject would have the same degree of cyanosis as an ordinary person with this quantity (5 g) of reduced hemoglobin in his blood. But in the polycythemic subject with a hemoglobin content of 200%, 4 5 volumes % is only 11% of the oxygen capacity of his blood, in the person with a normal hemoglobin content it constitutes 22 5%. The oxygen want in the former would be relatively slight as compared with that of the latter. Or put in another way and including the anemic subject—if the anoxemia were of the same degree in each, the anemic person would show little or no cyanosis, the polycythemic, a cyanosis of high degree and the normal subject a color intermediate in intensity. In a person with an abnormally high blood count a degree of slowing of the peripheral blood flow which would be without effect upon one possessing a normal hemoglobin content will result in cyanosis. Thus in regions such as the face, ear lobes, and hands, where the cutaneous vessels are well filled with blood, cyanosis is readily induced in the polycythemic subject by exposure to cold. In certain conditions associated with arterial anoxemia, e g , emphysema (p 6-73) and congenital heart disease, the red cell count is increased, and the cyanosis, in consequence, enhanced. It will be evident from these examples that the presence of cyanosis indicates anoxia but the absence of cyanosis does not rule out the possibility of anoxia.

THE FACTORS WHICH INFLUENCE THE DEPTH OF CYANOSIS CAUSED BY A GIVEN QUANTITY OF REDUCED HEMOGLOBIN

The State of Cutaneous Capillaries When these are dilated more of the dark-colored blood will be present in the skin than when they are constricted In the former instance cyanosis will of course be more pronounced Increased carbon dioxide tension in the peripheral blood causes capillary dilation, therefore, when retention of this gas accompanies oxygen want, as in venous congestion of superficial regions, etc the cyanosis is intense If the peripheral vessels are strongly constricted, as in shock, cyanosis may be inconspicuous

Pigmentation and Thickness of the Skin These factors obviously will modify the depth of the cyanotic color Cyanosis is more clearly evident in regions where the skin is thin and unpigmented The yellow discoloration of the skin caused by an excess of bilirubin in the blood (jaundice) tends to modify the cyanotic tint, but since the former stains the skin itself, while the discoloration due to reduced hemoglobin is confined to the capillary vessels, jaundice is likely to be just as intense in regions where the skin is thick as in those where it is thin The cyanotic discoloration can be temporarily abolished by pressure upon the skin whereas the icteric staining cannot Cyanosis does not appear in the conjunctivae but these are deeply colored in jaundice

Therapeutic Procedures· Oxygen Therapy

Oxygen administration is of great value in combating acute arterial anoxia as may occur in pneumonia, pulmonary edema, or obstruction to breathing, as well as in states such as congestive heart failure, or coronary thrombosis when, though the arterial blood contains the usual amount of oxygen, the tissues, owing to impairment of the circulation, suffer from a deficiency of oxygen

The chronic anoxia due to anemia, in which the hemoglobin is saturated with oxygen to the normal degree, is treated preferably by measures directed to the disease itself rather than by oxygen therapy In failure of the circulation, the inhalation of 100% oxygen will increase the oxygen content of the blood to revive a flagging respiratory center But the anoxia due to a shunt of blood from the right to the left side of the heart (p 6-76) or through a similarly completely unaerated portion of the lung will not be much benefited by oxygen inhalations, it is not possible to make the blood supplying healthy and well-aerated alveoli absorb any important amount of *extra* oxygen and so compensate for the shunted blood

When, however, the diffusion coefficient of oxygen (chap 6 1) through the alveolar membrane is reduced as a result of edema, thickening, or a coating of fluid, oxygen administration by raising the pressure of the gas in the alveolar air will increase its rate of diffusion across the pulmonary epithelium The oxygen saturation of the blood flowing through the damaged pulmonary tissue is increased

Consequently, in broncho- or lobar pneumonia when such changes are responsible for the anoxemia, in emphysema or in pulmonary edema whether from cardiac or pulmonary disease or from gas poisoning, the success of oxygen administration is often spectacular Anoxemia possibly promotes an increase in transudation of fluid into the alveoli In other words, a vicious circle may be established—edema inducing anoxemia and the latter increasing the edema—which is broken by oxygen administration

Anoxemia due to rapid *shallow breathing* is relieved by oxygen treatment since the oxygen tension of poorly ventilated alveoli is raised thereby The shallow breathing itself is likely to persist since it is primarily due to the local process in the lung acting upon the nerve endings rather than to the anoxemia Therefore, this type of breathing could, no more than the pulmonary lesion itself, be expected to be abolished by oxygen treatment

It has been said that when the arterial blood contains the normal quantity of oxygen, namely about 19 5 volumes $^{c}_{o}$ (saturation 97 5$^{c}_{c}$) as in the hypokinetic type of anoxia of congestive heart failure, oxygen inhalation cannot be of any great value, since the oxygen saturation can be raised only to 100%, and the total oxygen content increased by 2 2 volumes $^{c}_{c}$ (an increase of 1 7 volumes in simple solution together with 0 5 volumes combined with hemoglobin) That is, the total oxygen content of a person breathing 100% oxygen will be 21 7 volumes %. But this increase of 2 2 volumes $^{c}_{c}$

(representing an increase in O_2 content of 11%) is as pointed out by Comroe and Dripps (1950) of the highest importance in anoxia, for it raises very considerably the pressure at which oxygen is delivered to the tissues For example, when a patient is breathing ordinary air, every 100 ml of blood in passing through the capillaries loses 7 volumes of O_2 (normal about 5 vol), and, therefore, leaves the venous end of the capillary with a content of 12 5 volumes %, is 63% saturated, and has a Po_2 of around 32 mm Hg Now when 100% oxygen is breathed the extra 2 2 volumes % in the arterial blood is given up on entering the capillaries, which reduces the oxygen saturation only to that existing under ordinary conditions in the arterial blood, namely, 97 5% But the tissues owing to the slowing of the circulation abstract a further 4 8 volumes % to satisfy their need of 7 volumes % The venous blood, therefore, contains 14 7 volumes % of oxygen, and is over 73% saturated, instead of about 63% without oxygen administration, the Po_2 is nearly 40 mm Hg Thus the tissues are supplied with oxygen at a virtually normal head of pressure In congestive heart failure, therefore, especially if there is generalized edema, the administration of oxygen is often of great benefit

METHODS OF ADMINISTRATION

Oxygen is given usually by means of a nasal or oral-nasal catheter, by a specially designed face mask, or the patient is placed in a hood-tent or airtight cabinet in which the O_2 percentage is maintained at the required concentration Carbon dioxide (5 to 10%) is sometimes added, especially in CO poisoning, persistent hiccough and postoperative atelectasis, in order to encourage lung expansion In an oxygen tent the patient's own breath may raise the CO_2 to the range of 1 to 2%

In normal persons the breathing of 100% oxygen often causes a slight initial reduction of breathing of about 3% In anoxia, a much more pronounced depression or even complete cessation of breathing may occur This so-called *oxygen apnea* may be very alarming, especially during anesthesia But it is a sure indication of the existence of anoxia and the need for oxygen therapy Before oxygen administration the center had been driven by impulses from the peripheral chemoreceptors, the relief of the anoxia has abolished this source of stimulation In chronic anoxia, such as may be seen in congestive heart failure and pulmonary emphysema, especially when there is also carbon dioxide retention, the patient may pass into coma when oxygen is administered The unfavorable effects following administration of O_2 under these conditions are attributable to the further rise in Pco_2 which occurs as a result of the depression of breathing

Another effect of oxygen inhalation is the elimination of nitrogen A man of average weight (70 kg) breathing oxygen eliminates about 18 ml of nitrogen/min The rate of removal of the gas from all tissues is not however the same, it is removed most rapidly from the blood and the most vascular parts of the brain Pure oxygen is, therefore, administered to divers to prevent decompression sickness, and to pilots before high altitude flights

Helium (atomic weight 4, ⅛ density that of nitrogen) is lighter than any other gas except hydrogen Barach has applied this physical fact to reduce the respiratory effort in asthmatic attacks, in those suffering from laryngeal or tracheal obstruction and in certain other types of dyspnea A gas mixture is used in which helium is substituted for nitrogen (oxygen, 21%, helium, 79%) For the relief of anoxemia the oxygen percentage may be increased to 60 or 70%

Effects of Prolonged Breathing 0 5 to 1 Atmosphere of O_2 Ill effects other than the "oxygen apnea" described above may result from prolonged administration of oxygen Commonly subjects complain of substernal distress in 4 to 16 hr after beginning to breathe 100% O_2 at sea level The symptom usually persists for some hours after the patient resumes breathing of atmospheric air The pain evidently is related to tracheobronchitis, and it is not a problem unless oxygen tension in inspired air is in excess of 0 5 of an atmosphere Other manifestations of toxicity produced by breathing 70 to 100% O_2 at a pressure of 1 atm include fatigue, paresthesias in the hands and feet, joint pain, anorexia, nausea, and vomiting The cause of these symptoms are unknown (Comroe and Dripps, 1950)

In the ordinary clinical management of patients when oxygen is given in a tent or by nasal catheter oxygen toxicity does not become a factor since the O_2 content in inspired air is rarely, if ever, increased to above 50% Higher levels can be achieved if a close-fitting mask is used in the absence of rebreathing or if the subject is placed in an enclosure into which oxygen is introduced.

Effects of Oxygen Pressures Greater Than 1 Atmosphere By placing the patient in a pressure chamber it is possible to administer oxygen under a pressure of more than 1 atm Recently this method has been used for treatment of carbon monoxide poisoning, and its usefulness for other purposes is being investigated extensively (see Lambertsen, Bond, and Jacobsen, II, 1965) When oxygen is breathed at pressures around 2 to 4 atm severe symptoms indicating dysfunction of the higher central nervous system are experienced These include mood swings, indifference, loss of judgment, somnolence, irrational apprehension, and finally convulsions of the grand mal type The latter are followed by automatism and amnesia Vertigo also is a common symptom of oxygen poisoning The cause of the symptoms is not clear Reduction in the CO_2 carrying capacity of the blood, with a corresponding rise in CO_2 tension and cH in the tissues and venous blood, has been shown to occur in experimental animals, but it is not known if the symptoms in man are related entirely to these changes (Bean, 1945)

Mice die within 20 to 25 min when exposed to an oxygen tension of 8 atm , whereas a pressure of 8 atm of air is tolerated well

Effects of Breathing Air at Very High Pressures (Nitrogen Narcosis) The inert gases, including nitrogen, at certain pressures display the typical properties of anesthetic agents Whereas surgical anesthesia is produced in patients breathing 80% of an atmosphere of nitrous oxide or xenon, about 38 atm of nitrogen has been calculated to be necessary to have this effect In experimental animals a pressure of 30 atm of nitrogen is required to abolish righting reactions However, symptoms referrable to impairment of function of the central nervous system occur in an individual exposed to 6 atm of nitrogen (the equivalent of about 200 ft below the surface of water) Helium-oxygen mixtures have less anesthetic action than nitrogen-oxygen mixtures, hence divers breathing helium-oxygen can descend to much greater depths without the development of mental aberrations

Resuscitation

Until the last few years Schafer's prone pressure method of artificially inflating the lungs was most widely practiced But today either the *Holger-Nielson* or the *mouth-to-mouth method* is preferred

HOLGER-NIELSON OR BACK-PRESSURE-ARM-LIFT METHOD

After loosening the clothing around the neck and chest, drawing the tongue well forward, and wiping fluid, mucus, etc , from the mouth and throat in order to allow free passage of air, the subject is laid in the prone position (chest down) The arms are abducted at the shoulders to lie at right angles to the trunk and bent at the elbows (see Fig 6 40) The head is turned to one side with a cheek resting on the hands The operator places himself in front of the subject with one knee resting beside his head and near his forearm The foot of the operator's other side is planted near the subject's opposite forearm Then the operator places his hands, with fingers spread apart, upon the subject's back, thumbs touching, and on an imaginary line joining the axillae He then sways forward, bringing his weight to bear through his arms and hands upon the subject's back The operator should hold his arms straight at the elbows This movement compresses the chest and imitates natural expiration The operator then seizes the subject's arm above the elbows and draws them forward, at the same time as he himself straightens up This position which expands the chest and simulates natural inspiration is held for 2 sec He then swings forward again and repeats the first phase The double movement is repeated about 12 times a minute

MOUTH-TO-MOUTH METHOD

In a symposium published in the May, 1958 issue of the Journal of the American Medical Association it is stated that mouth-to-mouth resuscitation is superior to all manual methods for all age groups It is reported to be the only

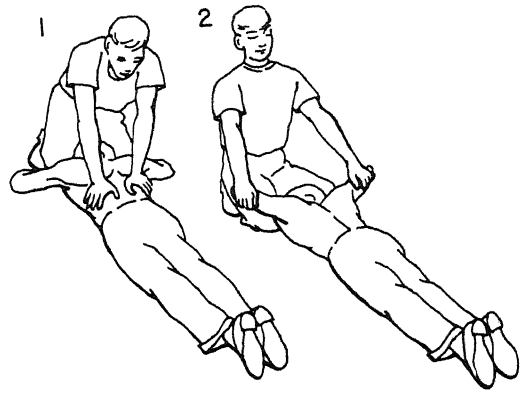


Fig 6 40 Illustrating the Holger-Nielsen method of resuscitation (From *First Aid Metropolitan Life Insurance Company*)

technique capable of producing adequate ventilation in all cases The single most important factor preventing adequate ventilation in this and some other methods, is mechanical obstruction of the airway by the tongue Proper extension of the neck and elevation of the jaw eliminates this obstruction To perform mouth-to-mouth resuscitation the subject is placed in the supine position with the head extended and the rescuer at the side of the head The rescuer grasps the lower jaw of the subject between his thumb and index finger and lifts it vertically upward The other thumb and index finger are used to clamp off the nostrils The rescuer then places his mouth over the subject's mouth and exhales into the airway a volume about two times the normal tidal volume (Fig 6 41) Volume and pressure required are easily judged by the rescuer He inflates the subject's lungs and thorax and notes the expansion of the chest The rescuer then removes his mouth from that of the subject to allow him to exhale The procedure should be performed about 12 to 20 times/min

The effectiveness and ease of performance of this type of resuscitation is improved by the use of an anesthesia face mask, or an oropharyngeal airway or an endotracheal tube

During recent years several methods have been developed for the long continued application of artificial respiration by various mechanical devices This work was begun by Thunberg in Lund, who devised an apparatus called the barospirator The subject was placed inside a metal chamber in which the pressure was raised and lowered rhythmically by means of the stroke of a large piston The interchange of air within the lungs was caused by a rise and fall of pressure of the air in the external atmosphere This apparatus was effective and a model was built large enough to accommodate patient, nurse, and doctor All three were ventilated without movement of the chest A certain amount of discomfort was experienced in this cabinet due to the change in pressure on the two sides of the eardrum during the increase and decrease of air pressure A more generally applicable model was produced by Drinker at Harvard In this type the patient's head remains outside the cabinet (Fig 6 42) The chest is expanded by reducing the pressure within the cabinet and as the pressure is raised again, the natural elasticity of the lung causes expiration Forced expiration, however, may be produced by raising the pressure above atmospheric Patients have been adequately ventilated with this apparatus for many months Some difficulty is occasionally encountered in the regulation of the rate and depth of respiration, but by determining the oxygen saturation of the arterial blood or watching for signs of cyanosis an observer can usually regulate the ventilation quite satisfactorily

Two other types of apparatus should be mentioned, *1*) the jacket model of Sahlin which operates like the Drinker machine but is applied only to the chest, and 2) the Bragg-Paul pulsator which consists of a hollow elastic bandage placed around the chest The bandage, when inflated by an electrically driven bellows, compresses the chest, which returns to the midposition by virtue of its own elasticity during the intervals between the compressions

The pulmotor and other similarly devised resuscitators, which force air into the chest and suck it out again as though the lungs were rubber bags, though so appealing in their mechanical efficiency, are physiologically unsound They are not only less effective than other methods but may be actually harmful The pulmonary tissue may be torn Those machines which inflate the lungs with an oxygen-carbon dioxide mixture, and permit expiration to take place simply by interrupting the air current and thus permitting collapse of the thorax, are sounder in principle, but precautions are necessary to insure that pressures produced in the lungs will not be excessive

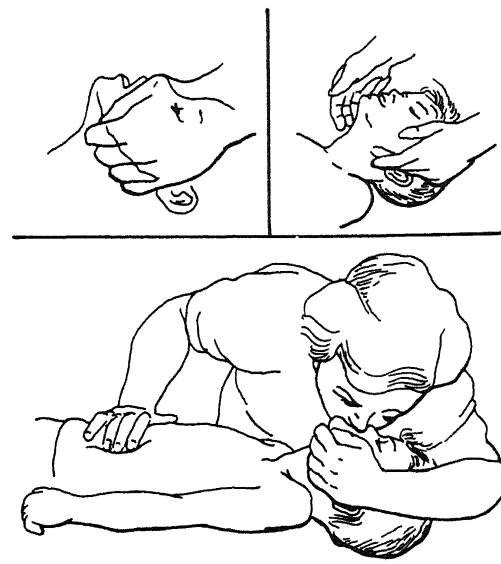


Fig 6 41 Mouth-to-mouth method of artificial respiration (Courtesy of Dr John Henderson and Duell Sloan, Pierce and Company Publishers New York)

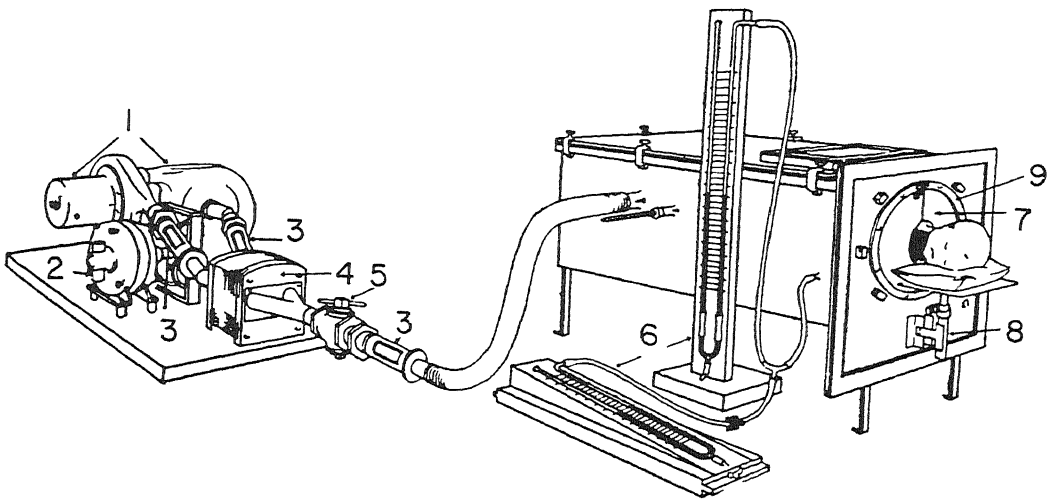


Fig 6 42 The Drinker respirator *1* pumps *2* motor *3* vents *4* alternate *5* valves *6* manometers *7* external shutters *8* adjustment for head rest *9* adjustable ring to hold collar in place

A mixture of oxygen and carbon dioxide (5% or 10% of the latter) commonly is used for resuscitation The high tension of CO_2 combined with the rhythmical inflation and deflation of the lungs, which presumably cause the discharge of afferent impulses, encourages the return of spontaneous breathing

RESUSCITATION OF THE NEWBORN

Asphyxia in the newborn infant is usually due to the failure of the lungs to expand fully, alveoli in areas of the pulmonary tissue remaining in the fetal state (atelectasis) The methods of resuscitation which may be effective in the adult are likely to fail Cutaneous stimulation, e g , slapping, hot or cold water, are time honored devices which are usually unsuccessful The object aimed at should be the expansion of the collapsed alveoli, and when the infant makes no respiratory effort this is best accomplished by the insufflation of the lungs with a carbon dioxide-oxygen mixture, carried out by the passage down the trachea of a tube, to the outer end of which a rubber bag filled with the gas mixture is attached When the breathing is not completely suspended but is weak, and the infant cyanosed, inhalations of a carbon dioxide-oxygen mixture are employed Also mouth-to-mouth resuscitation is quite effective in infants

BIBLIOGRAPHY

ACHARD O AND V BUCHER Courants d action bulbaires a rhythme respiratoire *Helv Physiol Pharmacol Acta* 12 265-283 1954

ADRIAN E D AND D W BRONK Impulses in single fibres of the phrenic nerve *J Physiol (London)* 66 81-101 1928

ADRIAN E D AND D W BRONK The discharge of impulses in motor nerve fibers *J Physiol (London)* 67 119 1929

ADRIAN E D AND F J J BUYTENDIJK Potential changes in the isolated brain stem of the goldfish *J Physiol (London)* 71 121-135 1931

ANICHKOV S V AND M L BELEN KIL *Pharmacology of the Carotid Body Chemoreceptors* Oxford Pergamon 1963

ASTRUP P A new approach to acid base metabolism *Clin Chem* 7 1 1961

AVERY M E *The Lung and Its Disorders in the Newborn Infant* Philadelphia Saunders 1964

AVERY M E AND J MEAD Surface properties in relation to atelectasis and hyaline membrane disease *J Dis Child* 97 517-523 1959

AVIADO D M JR AND C F SCHMIDT Reflexes from stretch receptors in blood vessels heart and lungs *Physiol Rev* 35 247-300 1955

BAKER S P AND F A HITCHCOCK Immediate effects of inhalation of 100% oxygen at one atmosphere on ventilation volume carbon dioxide output oxygen consumption and respiratory rate in man *J Appl Physiol* 10 363-366 1957

BALDWIN E DEF A COURNAND AND D W RICHARDS Pulmonary insufficiency I Physiological classification clinical methods of analysis standard values in normal subjects Medicine (*Balt*) 27 243-278 1948

BARACH, A L The therapeutic use of oxygen in heart disease *Ann Intern Med* 5 428-440 1931

BARACH A L AND D W RICHARDS JR Effects of treatment with oxygen in cardiac failure *Arch Intern Med (Chicago)* 48 325-347 1931

BARCLAY A E K J FRANKLIN AND M M PRICHARD *The Foetal Circulation and Cardiovascular System* Oxford Blackwell 1946

BARCROFT J *The Respiratory Functions of the Blood* London Cambridge Univ Press 1925

BARCROFT J AND E P POULTON The effect of carbonic acid on the dissociation curve of blood *J Physiol (London)* 46 IV 1913

BATES D V AND R V CHRISTIE *Respiratory Function in Disease* Philadelphia Saunders 1964

BEAN J W Effects of oxygen at increased pressure *Physiol Rev* 25 1-147 1945

BEHNKE A R Problems in the treatment of decompression sickness (and traumatic air embolism) *Ann N Y Acad Sci* 117 843-859 1965

BENESCH R E AND R BENESCH The reaction between diphosphoglycerate and hemoglobin *Fed Proc* 29 1101-1104 1970

BINGER C A L AND R L MOORE Changes in carbon dioxide tension and hydrogen ion concentration of the blood following multiple pulmonary embolism *J Exp Med* 45 633-641 1927

BISHOP B Reflex control of abdominal muscles during positive pressure breathing *J Appl Physiol* 19 224-232 1964

BJURSTEDT H Influence of the abdominal muscle tone on the circulatory response to positive pressure breathing in anesthetized dogs *Acta Physiol Scand* 29 145-162, 1953

BJURSTEDT H Respiratory system *Ann Rev Physiol* 19 151-174 1957

BLALOCK A Physiopathology and surgical treatment of congenital cardiovascular defects Harvey *Lect* 41 90-116 1945-1946

BLALOCK A AND H B TAUSSIG The surgical treatment of malformations of the heart in which there is pulmonary stenosis or pulmonary atresia *J A M A* 128 189-202 1945

BOREMA I W H BRUMMELKAMP AND N G MEIJNE (eds) *Clinical Applications of Hyperbaric Oxygen* Amsterdam Elsevier 1964

BOUVEROT P, R FLANDROIS R PUCCINELLI AND P DE JOURS Etude du role des chémorécepteurs artériels dans la régulation de la respiration pulmonaire chez le chien éveillé *Arch Int Pharmacodyn* 157 253-271, 1965

BRECHER G *Venous Return* New York Grune & Stratton Inc 1956

BRODIE D A AND H L BORISON Evidence for a medullary inspiratory pacemaker functional concept of central regulation of respiration *Amer J Physiol* 188 347-354 1957

BRONK D W AND L K FERGUSON The nervous control of intercostal respiration *Amer J Physiol* 110 700 1935

BROOKHART J M The respiratory effects of localized faradic stimulation of the medulla oblongata *Amer J Physiol* 129 709-723 1940

BROOKS C McC F F KAO AND B B LLOYD (eds) *Cerebrospinal Fluid and the Regulation of Ventilation* Oxford Blackwell 1965

BROWN I W AND B G COX *Proceedings of the Third International Conference on Hyperbaric Medicine* Washington D C National Academy of Sciences National Research Council 1966

BUNN H F M H MAY W F KOCHOLATY AND C F SHIELDS Hemoglobin function in stored blood *J Clin Invest* 48 311 1969

BURCHELL H B R L PARKER T J DRY E H WOOD J W PFNDER AND D G PUGH Cardiac catheterization in the diagnosis of various cardiac malformations and diseases *Mayo Clin Proc* 23 481-487 1948

BUYTENDIJK H J Oesophagusdruck en Longelasticiteit Thesis Groningen The Netherlands Univ of Groningen 1949

CAMPBELL E J M *The Respiratory Muscles and the Mechanics of Breathing* London Lloyd Luke 1958

CARLSON L D (conf chmn) Respiration In *Physiology in the Space Environment* Washington D C National Academy of Sciences National Research Council 1967 vol II pub 1485B

CARO C G (ed) *Advances in Respiratory Physiology* London Arnold 1966

CATCHPOLE H R AND I GERSH Pathogenetic factors and pathological consequences of decompression sickness *Physiol Rev* 27 360-397 1947

CATER D B S GARATTINI F MARINA AND I A SILVER Changes of oxygen tension in brain and somatic tissues induced by vasodilator and vasoconstrictor drugs *Proc Roy Soc London Ser B* 155 136 1961

CHARLTON G D READ AND J READ Continuous intra arterial P_O in normal man using a flexible microelectrode *J Appl Physiol* 18 1247 1963

CHEN K K C L ROSE AND G H A CLOWES Comparative values of several antidotes in cyanide poisoning *Amer J Med Sci* 188 767-781 1934

CHERNIAK R M AND L CHERNIAK *Respiration in Health and Disease* Philadelphia Saunders 1961

CHRISTENSEN H The chemical and nervous control of respiration In *Handbook of Respiratory Physiology* Randolph Field Texas United States Air Force School of Aviation Medicine 1954 p 103

CHRISTIANSEN J C G DOUGLAS AND J S HALDANE The absorption and dissociation of carbon dioxide by human blood *J Physiol (London)* 48 244 1914

CHRISTIE R V The elastic properties of the emphysematous lung and their clinical significance *J Clin Invest* 13 295-321 1934

CHRISTIE R V AND A L LOOMIS The pressure of aqueous vapor in the alveolar air *J Physiol (London)* 77 35 1932

CHRISTIE R V AND J C MEAKINS The intrapleural pressure in congestive heart failure and its clinical significance *J Clin Invest* 13 323-345 1934

CHURCHILL E D AND O COPE The rapid shallow breathing resulting from pulmonary congestion and edema *J Exp Med* 49 531-537 1929

CLEMENTS J A Surface tension of lung extracts *Proc Soc Exp Biol Med* 95 170-172 1957

COHEN M I Intrinsic periodicity of the pontile pneumotaxic mechanism *Amer J Physiol* 195 23-27 1958

COLE, P V AND L H HAWKINS The measurement of oxygen content of whole blood *Biomed Engin* 2 56 1967

COLLIER C R J E AFFELDT AND A F FARR Continuous rapid infrared CO_2 analysis *J Lab Clin Med* 45 526 1955

COMROE J H JR The hyperpnea of muscular exercise *Physiol*

Rev 24 319–339 1944
COMROE J H JR Measurements on arterial blood *Meth Med Res* 2 138 1950
COMROE J H JR *Physiology of Respiration* Chicago Year Book 1965
COMROE J H JR AND R D DRIPPS *The Physiological Basis of Oxygen Therapy* Amer Lecture Series Springfield Illinois Thomas 1950
COMROE J H JR R E FORSTER II A B DU BOIS W A BRISCOE AND E CARLSEN *The Lung* Ed 2 Chicago Year Book 1962
COMROE J H JR AND W S FOWLER Lung function studies *Amer J Med* 10 408 1951
COMROE J H JR AND L MORTIMER The respiratory and cardiovascular responses of temporally separated aortic and carotid bodies to cyanide nicotine phenyldiguanide and serotonin *J Pharmacol Exp Ther* 146 33–41 1964
COMROE J H JR AND C F SCHMIDT The part played by reflexes from the carotid body in the chemical regulation of respiration in the dog *Amer J Physiol* 121 75–97 1938
COMROE J H JR AND C F SCHMIDT Reflexes from the limbs as a factor in the hyperpnea of muscular exercise *Amer J Physiol* 138 536–547 1943
CONSOLAZIO C F R E JOHNSON AND L J PECORA *Physiological Measurements of Metabolic Functions in Man* New York McGraw Hill 1963
COOMBS H C AND F H PIKE The role of the afferent impulses in the control of respiratory movements *Amer J Physiol* 45 569 1917
CORDIER D AND C HEYMANS *Le Centre Respiratoire* Paris Hermann 1935
COURNAND A J S BALDWIN AND A HIMMELSTEIN *Cardiac Catheterization in Congenital Heart Disease* New York Commonwealth Fund 1949
COZINE R A AND S H NGAI Medullary surface chemoreceptors and regulation of respiration in the cat *J Appl Physiol* 22 117–121 1967
CROSFILL M L AND J G WIDDICOMBE Physical characteristics of the chest and lungs and the work of breathing in different mammalian species *J Physiol (London)* 158 1 1961
CUNNINGHAM D J C AND B B LLOYD (eds) *The Regulation of Human Respiration* Oxford Blackwell 1963
DARLING R C C A SMITH E ASMUSSEN AND F M COHEN Some properties of human fetal and maternal blood *J Clin Invest* 20 739 1941
DAVENPORT H W *The ABC of Acid Base Chemistry* Ed 4 Chicago Univ Chicago Press 1958
DAVIES P W AND D W BRONK Oxygen tension in mammalian brain *Fed Proc* 16 689 1957
DAWES G S AND J H COMROE JR Chemoreflexes from the heart and lungs *Physiol Rev* 34 167–201 1954
DAYMEN H Mechanics of air flow in health and in emphysema *J Clin Invest* 30 1175 1951
DEJOURS P *Respiration* New York Oxford Univ Press 1966
DE REUCK A V S AND M O'CONNOR *CIBA Foundation Symposium on Pulmonary Structure and Function* Boston Little Brown 1962
DICKENS F AND E NEIL (eds) Oxygen in the animal organism Oxford Pergamon 1964
DRIPPS R D AND J H COMROE JR The effect of the inhalation of high and low oxygen concentrations on respiration pulse rate ballistocardiogram and arterial oxygen saturation (oximeter) of normal individuals *Amer J Physiol* 149 277–291, 1947
DRIPPS R D AND J W SEVERINGHAUS General anesthesia and respiration *Physiol Rev* 35 741–777 1955
DU BOIS A B S Y BOTHELHO AND J H COMROE JR A new method for measuring airway resistance in man using a body plethysmograph values in normal subjects and in patients with pulmonary disease *J Clin Invest* 35 327 1956
DUNN J S The effects of multiple embolism of pulmonary arterioles *Quart J Med* 13 129–147 1920
EYSTER, J A E Clinical and experimental observations upon Cheyne Stokes breathing *J Exp Med* 8 565–613 1906
FYZAGUIRRE C AND H KOYANO Effects of hypoxia hypercapnia and pH on the chemoreceptor activity of the carotid body in vitro *J Physiol (London)* 178 385–409, 1965
EYZAGUIRRE C AND J LEWIN Chemoreceptor activity of the carotid body of the cat *J Physiol (London)* 159 222–237 1961
FAWCITT J J LIND AND C WEGELIUS The first breath *Acta Paediat Scand* 49 (Suppl 123) 5 1960
FENCL V T B MILLER AND J R PAPPENHEIMER Studies on the respiratory response to disturbances of acid base balance with deductions concerning the ionic composition of cerebral interstitial fluid *Amer J Physiol* 210 459–472 1966
FENN W O Mechanics of respiration *Amer J Med* 10 77 1951
FENN W O AND H RAHN (eds) *Handbook of Physiology* Washington D C *Amer Physiol Soc* 1964 vol I
FERGUSON J K W AND F J W ROUGHTON The direct chemical estimation of carbamino compounds of CO with haemoglobin *J Physiol (London)* 84 69 1934
FERRIS B G JR J MEAD AND J H OPIE Partitioning of respiratory flow resistance in man *J Appl Physiol* 19 653 1964
FERRIS B G JR AND D S POLLARD Effect of deep and quiet breathing on pulmonary compliance in man *J Clin Invest* 39 143 1960
FLOREZ J AND H L BORISON Tidal volume in CO regulation peripheral denervations and ablation of area postrema *Amer J Physiol* 212 985 991 1967
FLOURENS P Note sur le point vital de la moelle allongée *C R Acad Sci (Paris)* 33 437–439 1851
FLOURENS P Nouveaux details sur le noeud vital *C R Acad Sci (Paris)* 47 803–806 1858
FORSTER R E Exchange of gases between alveolar air and pulmonary capillary blood pulmonary diffusing capacity *Physiol Rev* 37 391 1957
FOWLER W S Intrapulmonary distribution of inspired gas *Physiol Rev* 1 (Suppl 2) 1 1952
FRANK N R J MEAD A A SIEBENS AND C F STOREY Measurements of pulmonary compliance in seventy healthy young adults *J Appl Physiol* 9 38 1956
FULTON J F (ed) *Decompression Sickness* Philadelphia Saunders 1951
GESELL R The chemical regulation of respiration *Physiol Rev* 5 551–595 1925
GESELL R Respiration and its adjustments *Ann Rev Physiol* 1 185–216 1937
GESELL R J BRICKER AND C MAGEE Structural and functional organization of the central mechanism controlling breathing *Amer J Physiol* 117 423–452 1936
GESELL R C S MAGEE AND J W BRICKER Activity patterns of the respiratory neurons and muscles *Amer J Physiol* 128 625–628 1940
GESELL R AND C MOYER The dual excitatory action of the vagal stretch reflex *Amer J Physiol* 131 674–680 1941
GLEICHMANN U D H INGVAR D W LUBBERS B K SIESJO AND G THEWS Tissue Po and Pco of the cerebral cortex related to blood gas tensions *Acta Physiol Scand* 55 127 1962
GRAY J S *Pulmonary Ventilation and Its Regulation* Springfield Illinois Thomas 1950
GRODINS F S Analysis of factors concerned in regulation of breathing in exercise *Physiol Rev* 30 220–239 1950
GRODINS F S Regulation of pulmonary ventilation *Physiologist* 7 319–333 1964
GROOM D E H WOOD H B BURCHELL AND R L PARKER The application of an oximeter for whole blood to diagnostic cardiac catheterization *Mayo Clin Proc* 23 601–609 1948
GUYTON A C J W CROWELL AND J W MOORE Basic oscillating mechanism of Cheyne Stokes breathing *Amer J Physiol* 187 395–398 1956
HABER E K W KOHN S H NGAI D A HOLADAY AND S C WANG Localization of spontaneous respiratory neuronal activities in the medulla oblongata of the cat a new location of the expiratory center *Amer J Physiol* 190 350–355 1957
HALDANE J S AND J G PRIESTLY *Respiration* Oxford Oxford Univ Press 1935
HARRISON T R *Failure of the Circulation* Baltimore Williams & Wilkins 1939
HARRISON T R J A CALHOUN G E CULLEN W E WILKINS AND C PILCHER Studies in congestive heart failure XV Reflex versus chemical factors in the production of rapid breathing *J Clin Invest* 11 133–154 1932

HARROP G Relation of the pulmonary gas diffusion constant to mountain sickness *J A M A* 80 1644 1923

HEEMSTRA H Respiration *Ann Rev Physiol* 18 121-144 1956

HENDERSON L J *Blood A Study in General Physiology* New Haven Conn Yale Univ Press 397 pp 1928

HENDERSON Y *Adventures in Respiration* Baltimore Williams & Wilkins 1938

HENDERSON Y AND M C HENDERSON The absorption of gas from any closed space within the body particularly in the production of atelectasis and after pneumothorax *Arch Intern Med (Chicago)* 49 88-93 1932

HERING, E, AND J BREUER Die Selbststeuerung der Athmung durch den Nervus vagus (abt II) Sitz Akad Wiss Wien 57 672-677 1868

HEYMANS C Role of the cardioaortic and carotid sinus nerves in the reflex control of the respiratory center *Neu Eng J Med* 219 157-159 1938

HEYMANS C AND J J BOUCKAERT *Le Sinus Carotidien et La Zone Homologue Cardioaortique* Paris Doin 1933

HEYMANS C AND E NEIL *Reflexogenic Areas of the Cardiovascular System* London Churchill 1958

HILL L *Caisson Sickness* London Arnold 1912

HOFF H E AND C G BRECKENRIDGE Intrinsic mechanisms in periodic breathing *Arch Neurol Psychiat* 72 11-42 1954

HUG E L'intoxication par acide cyanhydrique *C R Soc Biol (Paris)* 112 511-513 1933

HURTADO A Animals at high altitudes resident man In *Handbook of Physiology*, edited by D B DILL E F Adolph and C G Wilber Washington Amer Physiol Soc 1964 sect 4 chap 54

IVY A C High altitude problems in aviation *Fed Proc* 5 319-327 1946

JAEGER, M J AND A B OTIS Measurement of airway resistance with a volume displacement body plethysmograph *J Appl Physiol* 19 813 1964

JAMIESON, D AND H A S VAN DEN BRENK Effect of electrode dimensions on tissue Po_2 measurements in vivo *Nature (London)* 201 1227, 1964

KARLBERG P R B CHERRY F E ESCARDO AND G KOCH Respiratory studies in newborn infants II Pulmonary ventilation and mechanics of breathing in the first minute of life including the onset of respiration *Acta Paediat Scand* 51 121 1962

KETY S Determinants of tissue oxygen tension *Fed Proc* 16 666 1957

KETY S S AND C F SCHMIDT The effects of altered arterial tensions of carbon dioxide and oxygen on cerebral blood flow and cerebral oxygen consumption of normal young men *J Clin Invest* 27 484-492, 1948

KOUNTZ, W B AND H L ALEXANDER Emphysema *Medicine (Balt)* 13 251-316, 1934

KROGH, A *The Comparative Physiology of Respiratory Mechanisms* Philadelphia Univ Pennsylvania Press 1941

KROGH, M Diffusion of gases through the lungs of man *J Physiol (London)* 49 271-300, 1914-1915

LAMBERTSEN, C J The philosophy of extremes for the gaseous environment of manned, closed ecological systems *Aerospace Med* 34 291-299, 1963

LAMBERTSEN, C J Physiological interactions and gaseous environment in manned exploration of space *Fed Proc* 22 1046-1050, 1963

LAMBERTSEN C J Effects of oxygen at high partial pressure In *Respiration Handbook of Physiology* edited by W O Fenn and H Rahn Washington Amer Physiol Soc 1965a

LAMBERTSEN C J Therapeutic gases, oxygen, carbon dioxide and helium In *Drill's Pharmacology in Medicine*, Ed 3, edited by Di Palma New York McGraw Hill, Inc 1965b

LAMBERTSEN, C J Part V Respiration In *Medical Physiology*, edited by V B Mountcastle St Louis Mosby 1968 p 613

LAMBERTSEN, C J, G BOND, AND J H JACOBSEN, II (eds) Hyperbaric oxygenation *Ann N Y Acad Sci* 117 647-890, 1965

LEGALLOIS, C J J Oeuvres (avec des notes de M Pariset) Paris Le Rouge 1824, 2 vols pp 363, 402

LEITNER, L -M, B PAGES, R PUCCINELLI AND P DEJOURS Etude simultanée de la ventilation et des décharges des chémorécepteurs du glomus carotidien chez le chat I Au cours d inhalations brèves d oxygene pur *Arch Int Pharmacodyn* 154 421-426 1965

LENFANT C AND C AUCUTT Measurement of blood gases by gas chromatography *Resp Physiol* 1 398 1966

LENFANT C J D TORRANCE R D WOODSON P JACOBS AND C A FINCH Role of organic phosphates in the adaptation of man to hypoxia *Fed Proc* 29 1115-1117 1970

LEUSEN I R Chemosensitivity of the respiratory center influence of CO_2 in the cerebral ventricles on respiration *Amer J Physiol* 176 39-44 1954a

LEUSEN I R Chemosensitivity of the respiratory center influence of changes in the H and total buffer concentrations in the cerebral ventricles on respiration *Amer J Physiol* 176 45-51 1954b

LEUSEN J The control of respiration in health and disease In *The Biological Basis of Medicine* edited by E E Bittar and N Bittar New York Academic Press 1969 vol 6 p 289

LILJESTRAND Å Respiratory reactions elicited from medulla oblongata of the cat *Acta Physiol Scand* 29(Suppl 106) 321-393 1953

LILLIENTHAL J L JR, R L RILEY D D PROEMMEL AND R E FRANKE An experimental analysis in man of the oxygen pressure gradient from alveolar air to arterial blood during rest and exercise at sea level and at altitude *Amer J Physiol* 147 199 1946

LOESCHCKE G C Spielen fur die Ruheatmung des Menschen vom O_2 Druck abhangige Erregungen der Chemoreceptoren eine Rolle? *Pfluger Arch* 257 349-362 1953

LOESCHKE G C *Cerebrospinal Fluid and the Regulation of Ventilation* edited by C McC Brooks F F Kao and B B Lloyd Oxford Blackwell 1965 p 193

LUFT U L Physiological aspects of pressure cabins and rapid decompression In *Handbook of Respiratory Physiology* Randolph Field Texas United States Air Force School of Aviation Medicine 1954 p 129

LUMSDEN T The regulation of respiration *J Physiol (London)* 58 part I, 81-91 part II 111-126 1923

LUNDSGAARD C AND D D VAN SLYKE Cyanosis Medicine (*Balt*) 2 1-76 1923

LYSYJ I AND P R NEWTON Determination of oxygen nitrogen and carbon dioxide in air samples *J Chromatogr* 11 173 1963

MAGOUN, H W (ed) *Handbook of Physiology* Washington, D C Amer Physiol Soc 1960, sect 1 vol II

MARCKWALD, M Die Bedeutung des Mittelhirns fur die Atmung *Z Biol* 26 259-289 1890

MARSHALL, R AND A B DU BOIS Measurements of the viscous resistance of the lung tissues in normal man *Clin Sci* 15 161 1946

McILROY, W B, J MEAD N J SILVERSTONE AND E P RADFORD JR Measurement of lung tissue viscous resistance using gases of equal kinematic viscosity *J Appl Physiol* 7 485 1955

MEAD J Measurement of inertia of the lungs at increased ambient pressure *J Appl Physiol* 9 208 1956

MEAD J Volume displacement plethysmograph for respiratory measurements in human subjects *J Appl Physiol* 15 736 1960

MEAD, J Mechanical properties of lungs *Physiol Rev* 41 281, 1961

MEAD J 1962 Mechanics of respiratory structures In *Ciba Foundation Symposium on Pulmonary Structure and Function*, edited by A V S de Reuck and M O'Connor Boston Little, Brown 1962 p 111

MEAKINS J C Dyspnea *J A M A* 103 1442-1445, 1934

MEAKINS, J AND H W DAVIES Observations on the gases in human arterial and venous blood *J Path Bact* 23 451-461 1920

MEAKINS J AND H W DAVIES Respiratory function in disease Edinburgh Oliver, 1925

MEANS J H Dyspnea Medicine (*Balt*) 3 309, 1924

METZ B Brain acetylcholinesterase and a respiratory reflex *Amer J Physiol* 192 101-105 1958

MEYER, J S, AND F GOTOH Interaction of cerebral hemodynamics and metabolism *Neurology* 11, no 4 part 2, p 46, 1961

MILLIKAN, G A *Handbook of Respiration Data in Aviation* Washington, D C National Research Council, Philadelphia Saunders, 1958

MITCHELL R A C R BAINTON, J W SEVERINGHAUS AND G EDELIST Respiratory response and CSF pH during disturbances in blood acid base balance in awake dogs with denervated aortic and carotid bodies (abstr) *Physiologist* 7 208 1964

MITCHELL, R A, H H LOESCHCKE, J W SEVERINGHAUS,

B W RICHARDSON AND W H MASSION Regions of respiratory chemosensitivity on the surface of the medulla *Ann N Y Acad Sci* 109 661–681 1963

MONNIER M *Functions of the Nervous System* New York Elsevier 1968 p 350

NATELSON S *Microtechniques of Clinical Chemistry* Springfield Illinois Thomas 1957

NEERGAARD K v AND K WIRZ Uber eine Methode fur Messung der Lungen elastizität am lebenden Menschen insbesondere beim Emphysem *Z Klin Med* 105 35 1927

NEERGAARD K v AND K WIRZ Die Messung der Stromungswiderstände in den Atemwegen des Menschen insbesondere bei Asthma und Emphysem *Z Klin Med* 105 51 1927

NGAI S H AND S C WANG Organization of central respiratory mechanisms in the brain stem of the cat localization by stimulation and destruction *Amer J Physiol* 190 343–349 1957

NIELSEN M AND H SMITH Studies on the regulation of respiration in acute hypoxia (With an appendix on respiratory control during prolonged hypoxia) *Acta Physiol Scand* 24 293–313 1952

O LEARY J L AND L A COBEN The reticular core—1957 *Physiol Rev* 38 243–276 1958

OTIS A B The work of breathing *Physiol Rev* 34 449 1954

OTIS A B Respiration *Ann Rev Physiol* 20 159–178 1958

OTIS A B The control of respiratory gas exchange between blood and tissues In *The Regulation of Human Respiration*, edited by D J C Cunningham and B B Lloyd Oxford Blackwell 1963, p 111

OTIS A B Quantitative relationships in steady state gas exchange In *Handbook of Physiology* edited by W O Fenn and H Rahn Washington D C American Physiological Society 1964 vol I p 681

PAINTAL A S Vagal afferent fibers *Ergebn Physiol* 52 74–156 1963

PAPPENHEIMER J R V FENCL, S R HEISEY AND D HELD Role of cerebral fluids in control of respiration as studied in unanesthetized goats *Amer J Physiol* 208 436–450, 1965

PARTRIDGE, G Vagal and phrenic impulses and respiration *Canad Med Ass J* 33 11–22 1935

PATTLE R E The lining layer of the lung alveoli *Brit Med Bull* 19 41 1963

PAYNE J P AND D W HILL (eds) *A Symposium on Oxygen Measurements in Blood and Tissues and Their Significance* London Churchill, 1966

PEMBREY M S AND R W ALLEN Observations upon Cheyne Stokes respiration *J Physiol (London)* 32 xviii 1905

PETERS, J P, AND D D VAN SLYKE *Quantitative Clinical Chemistry* Baltimore Williams & Wilkins, 1932 vol II

PI SUÑER A The regulation of the respiratory movement by peripheral chemo receptors *Physiol Rev* 27 1–38, 1947

PITTS R F Organization of the respiratory center *Physiol Rev* 26 609–630 1946

PITTS R F H W MAGOUN, AND S W RANSON Localization of the medullary respiratory centers in the cat *Amer J Physiol* 126 673–688, 1939a

PITTS, R F H W MAGOUN AND S W RANSON Interrelations of the respiratory centers in the cat *Amer J Physiol* 126 689–707 1939b

PITTS R F, H W MAGOUN AND S W RANSON The origin of respiratory rhythmicity *Amer J Physiol* 127 654–670, 1939

PORTER, R (ed) Breathing Hering Breuer Centenary Symposium London Churchill, 402 pp 1970

PORTER, W T The path of the respiratory impulse from the bulb to the phrenic nuclei *J Physiol (London)* 17 455–485, 1895

PUGH, L G C E Resting ventilation and alveolar air on Mount Everest with remarks on the relation of barometric pressure to altitude in mountains *J Physiol (London)* 135 590–610, 1957

RADFORD E P, JR Recent studies of mechanical properties of mammalian lungs In *Tissue Elasticity* edited by J W Remington Washington, D C Amer Physiol Soc 1957, p 177 and 185

RADFORD, E P, JR The physics of gases In *Handbook of Physiology*, edited by W O Fenn and H Rahn Washington, D C Amer Physiol Soc 1964, sect 3, vol I p 125

RAHN H AND L FARHI Ventilation, perfusion, and gas exchange In *Handbook of Physiology*, edited by W O Fenn and H Rahn Washington D C Amer Physiol Soc 1964, sect 3 vol I p 735

RAHN H, AND W O FENN *A Graphical Analysis of the Respiratory Gas Exchange* Washington, D C Amer Physiol Soc, 1955

RAHN H AND A B OTIS Man's respiratory response during and after acclimatization to high altitude *Amer J Physiol* 157 445–462 1949a

RAHN H AND A B OTIS Continuous analysis of alveolar gas composition during work hyperpnea hypercapnia and anoxia *J Appl Physiol* 1 717 1949b

RAHN H OTIS A B CHADWICK L E AND FENN W O Pressure volume diagram of thorax and lung *Amer J Physiol* 146 161–178 1946

RILEY R L AND A COURNAND Ideal alveolar air and the analysis of ventilation perfusion relationships in the lungs *J Appl Physiol* 1 825 1949

RILEY R L AND A COURNAND Analysis of factors affecting partial pressures of oxygen and carbon dioxide in gas and blood of lungs theory *J Appl Physiol* 4 77 1951

RILEY R L A COURNAND AND K W DONALD Analysis of factors affecting partial pressures of oxygen and carbon dioxide in gas and blood of lungs methods *J Appl Physiol* 4 102 1951

RILEY R L J L LILIENTHAL JR D D PROEMMEL AND R E FRANKE On the determination of the physiologically effective pressures of oxygen and carbon dioxide in alveolar air *Amer J Physiol* 147 191 1946

RILEY R L A B OTIS AND C S HOUSTON Respiratory features of acclimatization to altitude In *Handbook of Respiratory Physiology* Randolph Field Texas United States Air Force School of Aviation Medicine 1954 p 143

ROHRER F K NAKESONE AND K WIRZ Physiologie der Atembewegung In *Handbuch der Normalen und Pathologischen Physiologie II* Berlin Springer 1925, p 70

ROSE Z B Enzymes controlling 2 3 diphosphoglycerate in human erythrocytes *Fed Proc* 29 1105–1111 1970

ROSSI G F AND A ZANCHETTI The brain stem reticular formation anatomy and physiology *Arch Ital Biol* 115 263–265 1957

ROSSI BERNARDI L AND F J W ROUGHTON The specific influence of carbon dioxide and carbamate compounds on the buffer power and Bohr effects in human haemoglobin solutions *J Physiol (London)* 189 1–29 1967

ROSSIER P H A A BUHLMANN AND K WEISINGER *Respiration* edited and translated by P C Luchsinger and K M Moser St Louis Mosby 1960

ROUGHTON F J W Some recent work on the chemistry of carbon dioxide transport by the blood *Harley Lect* 39 96 1943–1944

ROUGHTON F J W The average time spent by the blood in the human lung capillary and its relation to the rates of CO uptake and elimination in man *Amer J Physiol* 143 621, 1945

ROUGHTON F J W Transport of oxygen and carbon dioxide In *Handbook of Physiology* edited by W O Fenn and H Rahn Washington D C Amer Physiol Soc 1964 sect 3 vol I p 767

ROWE R D AND A MEHRIZI The neonate with congenital heart disease Philadelphia Saunders 1968

SAHLIN B Uber den antagonismus Zwischen methylenblau und cyankalium (Concerning the antagonism between methylene blue and potassium cyanide) *Skand Arch Physiol* 47 284–291 1926

SALAZER E AND J H KNOWLES An analysis of pressure volume characteristics of the lungs *J Appl Physiol* 19 97 1964

SCHLAFKE M S, AND H H LOESCHCKE Lokalisation eines an der Regulation von Atmung und Kreislauf beteiligten Gebietes an der ventralen Oberflache der Medulla oblongata durch Kalte blockade *Pfluger Arch* 297 201–220 1968

SCHMIDT, C F, AND J H COMROE JR Functions of the carotid and aortic bodies *Physiol Rev* 20 115–157, 1940

SCHMIDT C F AND J H COMROE, JR Respiration *Ann Rev Physiol* III 151–184, 1941

SCHNEIDER E C Physiological effects of altitude *Physiol Rev* I 631–659, 1921

SCHOPP, R T, T M GILFOIL, AND W B YOUMANS Mechanisms of respiratory stimulation during hemorrhage *Amer J Physiol* 189 117–122, 1957a

SCHOPP, R T, T M GILFOIL AND W B YOUMANS Mechanisms of respiratory responses to vasodilator drugs Urecholine (carbaminoyl beta methylcholine) and nitroglycerine *Amer J Physiol* 189 123–128, 1957b

SCOTT R W Observations on the pathologic physiology of chronic pulmonary emphysea *Ann Intern Med* 26 544–560 1920

SEVERINGHAUS, J W Oxyhemoglobin dissociation curve correction

for temperature and pH variation in human blood *J Appl Physiol* 12 485 1958

SEVERINGHAUS J W AND A F BRADLEY Electrodes for blood Po and Pco_2 determination *J Appl Physiol* 13 515 1958

SEVERINGHAUS J W M STUPFEL AND A F BRADLEY Variations of serum carbonic acid pK with pH and temperature *J Appl Physiol* 9 197 1956

SIGGAARD ANDERSON, O The first dissociation exponent of carbonic acid as a function of pH *Scand J Clin Lab Invest* 14 587 1962a

SIGGAARD ANDERSON O The pH log Pco_2 blood acid base nomogram *Scand J Clin Lab Invest* 14 598, 1962b

SIGGAARD ANDERSON O AND K ENGEL A new acid base nomogram *Scand J Clin Lab Invest* 12 177 1960

SINGER, R B AND A B HASTINGS An improved clinical method for the estimation of disturbances of the acid-base balance of human blood Medicine (*Balt*) 27 223 1948

SLONIM N F AND J L CHAPIN *Respiratory Physiology* St Louis Mosby, 1967

SMITH C A *The Physiology of the Newborn Infant* Springfield, Illinois Charles C Thomas 1959

SMITH G Carbon monoxide poisoning *Ann N Y Acad Sci* 117 684-687 1965

STADIE W C The treatment of anoxemia in pneumonia in an oxygen chamber *J Exp Med* 35 337-360, 1922

STELLA G On the mechanism of production and the physiological significance of 'apneusis" *J Physiol (London)* 93 10-23, 1938a

STELLA G The dependence of the activity of the "apneustic centre on the carbon dioxide of the arterial blood *J Physiol (London)* 93 263-275, 1938b

STONE, D J Applied pulmonary physiology *Advances Intern Med* VII 243-283, 1955

Symposium on Respiratory Physiology in Manned Spacecraft *Fed Proc* 22 1022-1063, 1963

TORRES G E Validation of oxygen electrode for blood *J Appl Physiol* 18 1008 1963

TUCKER, V A Method for oxygen content and dissociation curves on microliter blood samples *J Appl Physiol* 23 410, 1967

TURNER, J K Effects of right and left cardiac distension on steady state abdominal contractions (abstr) *Fed Proc* 20 120, 1961

TYUMA I AND K SHIMIZU Effect of organic phosphates on the difference in oxygen affinity between fetal and adult human hemoglobin *Fed Proc* 29 1112-1114, 1970

VAN LIERE, E J *Anoxia Its Effect on the Body* Chicago Univ Chicago Press 1942

VAN SLYKE D D, R T DILLON AND R MARGARIA Studies of gas and electrolyte equilibria in blood solubility and physical state of atmospheric nitrogen in blood cells and plasma *J Biol Chem* 105 571 1934

VAN SLYKE D D AND J PLAZIN *Micromanometric Analyses* Baltimore Williams & Wilkins 1961

VAN SLYKE D D AND J SENDROY Studies of gas and electrolyte equilibria in blood XV Line charts for graphic calculations by the Henderson Hasselbalch equation and for calculating plasma carbon dioxide content from whole blood content *J Biol Chem* 79 781 1928

WANG S C S H NGAI AND M J FRUMIN Organization of central respiratory mechanisms in the brain stem of the cat genesis of normal respiratory rhythmicity *Amer J Physiol* 190 333-342 1957

WENDELL W B Methylene blue and cyanide poisoning *J A M A* 100 1054-1055 1933

WEST J B *Ventilation/Blood Flow and Gas Exchange* Oxford Blackwell, 1965

WHITEHORN W V A LEIN, AND A EDELMANN The general tolerance and cardiovascular responses of animals to explosive decompression *Amer J Physiol* 147 289-298, 1946

WHITEHORN W V A LEIN A EDELMANN AND F A HITCHCOCK The effect of explosive decompression on cerebrospinal fluid pressure *Amer J Physiol* 148 253-258 1947

WHITEHORN W V A LEIN, AND F A HITCHCOCK The effect of binding the abdomen and thorax on pulmonary lesions produced by explosive decompression *J Aviat Med* 18 102-104, 1947

WHITTENBERGER, J L Artificial respiration *Physiol Rev* 35 611-628 1955

WINTERSTEIN H Die chemische Steuerung der Atmung *Ergebn Physiol* 48 328-528, 1955

WOOD E H, W F SUTTERER AND L CRONIN Oximetry In *Medical Physics*, edited by O Glasser Chicago Year Book, 1960 vol 3 p 416

WYSS, O A M Respiratory centre and reflex control of breathing *Helv Physiol Pharmacol Acta* (Suppl 10) 5-35 1954

WYSS, O Die nervose Steureung der Atmung *Ergebn Physiol* 54 1-479 1964

YOUMANS W B Editorial role of respiratory responses to changes in arterial blood pressure *Anesthesiology* 19 552-554 1958

YOUMANS, W B Q R MURPHY J K TURNER L D DAVIS D I BRIGGS AND A S HOYE Activity of abdominal muscles elicited from the circulatory system *Amer J Phys Med* 42 1-70, 1963

YOUMANS W B AND R T SCHOPP Cheyne-Stokes breathing after denervation of carotid and aortic chemoreceptors and sino aortic pressoreceptors *Proc Soc Exp Biol Med* 95 100-101, 1957

7

Endocrine Control Systems

1 / INTRODUCTION TO ENDOCRINE CONTROL SYSTEMS

Signals that allow an animal to respond to changes in its environment are transmitted either by spread of depolarization through nerves or by release of hormones into the blood A hormone is a substance that is produced by a gland or specialized neurosecretory cell, is often stored in this cell, and is released into the blood in response to some specific stimulus Release may be triggered either by a nerve impulse or by a change in the concentration of some substance in the blood which is perfusing the gland (Fig 7 1) When the hormone reaches its target cells, it modifies the function of the cell to remove the stimulus that originally caused release of the hormone This "negative-feedback" system operates to maintain homeostasis within the organism Signals transmitted through nerves have the advantage of speed and localization to single cells within a large group of similar cells On the other hand, a hormone affects cells of varying type over longer times

The major endocrine glands and hormones that are of importance in controlling body function will be discussed in this section These include

Hypothalamus thyrotropin-releasing factor (TRF), corticotropin-releasing factor (CRF), growth hormone-releasing factor (GRF), follicle-stimulating hormone-releasing factor (FRF), leutinizing hormone-releasing factor (LRF), and prolactin-inhibitory factor (PIF)

Hypophysis thyrotropin (TSH), adrenocorticotropic hormone (ACTH), growth hormone (GH), follicle-stimulating hormone (FSH), leutinizing hormone (LH), prolactin, melanocyte-stimulating hormone (MSH), vasopressin, and oxytocin

Thyroid thyroxine, triiodothyronine, and calcitonin

Parathyroid parathyroid hormone

Pancreatic islets insulin and glucagon

Adrenal glucocorticoids, mineralocorticoids, androgens, epinephrine, and norepinephrine

Ovary estrogen and progesterone

Testes testosterone

Skin vitamin D

Hormones involved in control of the gastrointestinal tract, cardiovascular system, and kidney are discussed in other sections of this text

METHODS OF IDENTIFYING ENDOCRINE GLANDS AND HORMONES

Removal of a gland that is suspected of endocrine function is often an initial approach to defining its function For example, when the anterior lobe of the pituitary is removed, there are important abnormalities in the animal These are *a*) arrested growth, *b*) atrophy of the gonads, and indirectly of the accessory organs of sex, *c*) suppression of milk secretion and involution of the mammary glands, *d*) atrophy of the thyroid and adrenal cortex, *e*) lowered metabolic rate, *f*) a tendency to hypoglycemia and reduction in liver and muscle glycogen, and *g*) depression of spontaneous activity and diminished resistance to infection and shock Since this gland secretes hormones that control other endocrine glands, effects of hypophysectomy are widespread This example emphasizes the range of functions that depend upon endocrine function

The second step in identifying the function of an endocrine gland is to make an extract of the tissue and to administer the extract to a hormone-deficient animal in the hope that it will correct the abnormalities that are present One of the classical demonstrations of the effect of an extract on such an abnormality was provided by F G Banting and C H Best, one of the original editors of this textbook When the pancreas is removed from dogs, the glucose content of the blood begins to rise within a short time, the urine gives a positive test for sugar, nitrogen excretion is elevated, and death occurs within 2 or 3 weeks Banting and Best (1922) were the first to obtain an extract

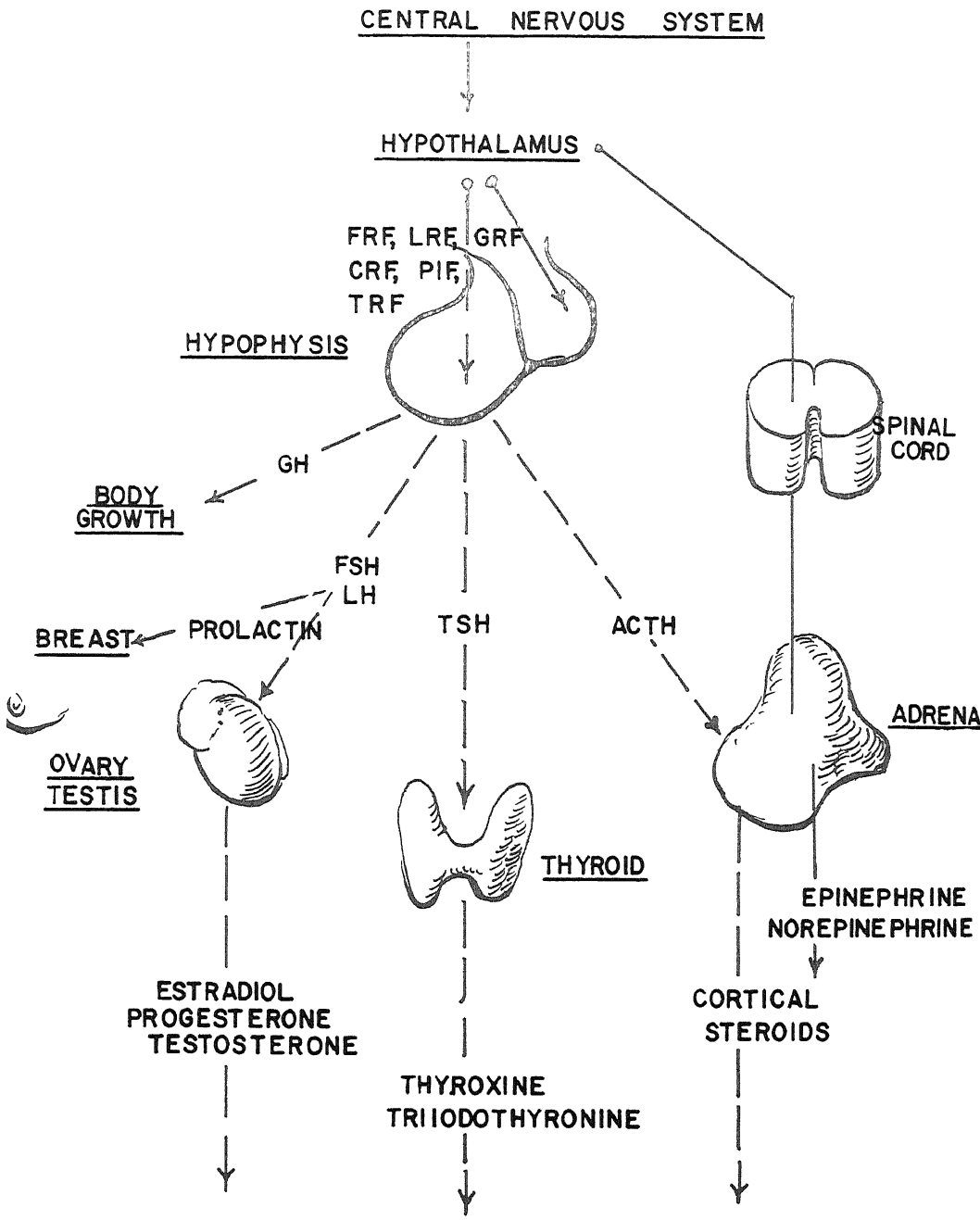


Fig 7 1 The relationship between the nervous system and the endocrine system ——nerve pathway ---- -- vascular pathway Note that neurosecretory structures the hypothalamus neurohypophysis and adrenal medulla are regulated by a direct nerve supply whereas the activity of other glands is controlled by humoral or hormonal agents carried in the blood

of the pancreas that would reverse these abnormalities of diabetes in the completely depancreatized dog (Fig 7 2) Subsequently, Abel et al (1927) crystallized insulin from extracts of pancreatic tissue

Once a hormonal activity has been identified in an extract of a gland, the substance responsible for the activity must be purified and the chemical structure determined Since many of the hormones are peptides or proteins, struc-

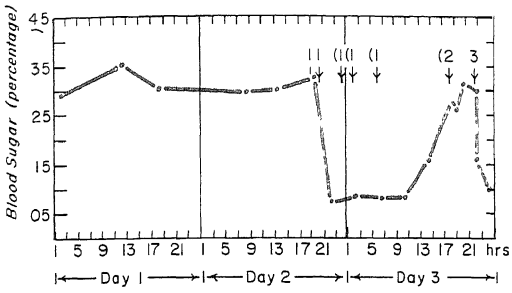


Fig 7 2 Effect of insulin of the blood sugar curve of a depancreatized dog *1*) Injection of extract of degenerated pancreas *2*) extract after incubation with pancreatic juice and *3*) extract incubated without pancreatic juice (Redrawn from Banting and Best 1922)

tures of this class of hormones have been determined only in the past 20 years Structures of the steroid hormones yielded more rapidly to chemical analysis and these structures have been determined during the past 50 years In these regards, endocrinology is a science of the 20th Century since E H Starling first used the word "hormone" in 1905 in referring to the activity of secretin, a gastrointestinal hormone

SYNTHESIS AND RELEASE OF HORMONES

Endocrine glands have certain anatomical features in common *1*) They are typical glands (except for the hypothalamus and neurohypophysis which belong to a special class of neurosecretory organs) in that they are composed of masses of secretory cells arranged in columns or clumps These cells may occasionally, as in the case of the thyroid, be arranged in alveoli or acini with their secretory product stored in the lumen More commonly, however, the hormones or their parent substances are stored as granules or droplets in the cells *2*) Although glandular in nature, no ducts are present *3*) Since the blood functions as a "duct" and carries off secretory products, they are richly vascularized

Protein hormones are synthesized on ribosomes that are attached to the endoplasmic reticulum The hormone is packaged into storage or "zymogen" granules in the Golgi apparatus The type of granule that is formed varies from one cell to another and serves as a means of identifying the hormone that is in the granule Release of hormone involves a process known as *emiocytosis* or reverse pinocytosis The membrane of the storage granules fuses with the plasma membrane of the cell and the product is released into the interstitial space

Steroid hormones are synthesized from cholesterol by a complicated series of reactions that occur in both the cytoplasm and mitochondria of these cells These cells are particularly rich in mitochondria and in some instances adaptations of mitochondrial structure to facilitate steroid synthesis are present In these glands the substrate for steroid synthesis, cholesterol is stored in lipid droplets as cholesterol esters When the gland is stimulated to produce hormone, it is synthesized either from this stored precursor or from cholesterol that is present in the plasma

ASSAY OF HORMONES

Hormones have effects when present in extremely low concentrations As a result, assay methods that are sufficiently sensitive and specific have been difficult to develop Initially, all hormones were measured by bioassays that were dependent upon the effect of a hormone on either a living animal or tissue preparation These assays are often insensitive and nonspecific When the structure of some hormones was determined, chemical assays were devised Thyroid hormone and adrenal and gonadal steroids are satisfactorily estimated by these methods

A particularly difficult problem is the estimation of peptide hormones in the concentrations normally found in plasma, millimicrogram per milliliter Since other proteins are present in the range of milligrams per milliliter, it is not possible to purify and quantitate peptide hormones by the usual techniques of fractionation A special method of assay was devised by Berson and Yalow (1964) and is known as a radioimmunoassay In order to use this method, a pure hormone must be prepared by the gland which synthesizes the compound This hormone is injected into an appropriate species of animal Since the animal reacts to the hormone in the same manner as it does to other foreign protein, antibodies against the hormone are formed After repeated injections of hormone are given, the serum of the animal contains a high titer of antibody against the purified hormone This serum can be used to measure the immunoreactive hormone in plasma and other body fluids It should be

stressed that immunological reactivity rather than biological activity is measured If inactive fragments of hormone react with the antibody or if hormone is bound to components in the plasma that prevent its interaction with antibody, these activities may not be the same

Once a potent antibody to a pure hormone is available, it can be used to measure plasma levels of peptide hormones An example of the use of this method to measure plasma levels of growth hormone is shown in Figures 7 3–7 5 The basic principle is to set up competitive reactions between radioactively-labeled hormone and unlabeled hormone for this specific antibody (Fig 7 3) Radioactive hormone is prepared by reaction of a portion of the tyrosine residues of the protein with ^{131}I After incubation of the mixture of labeled and unlabeled hormone with the antibody, labeled hormone which is antibody-bound is separated from free hormone by electrophoresis (Fig 7 4) The fraction of labeled hormone that is antibody-bound decreases as more unlabeled hormone is added to compete for the antibody

Free labeled-HGH (F)	Unlabeled-HGH (*in known standards or in body fluids*)
+	+
specific antibody	specific antibody
↓↑	↑↓
Antibody-bound labeled-HGH (B)	Antibody-bound unlabeled-HGH

Fig 7 3 Competing reactions that form the basis of the radioimmunoassay of human growth hormone (HGH) (From Yalow and Berson 1968)

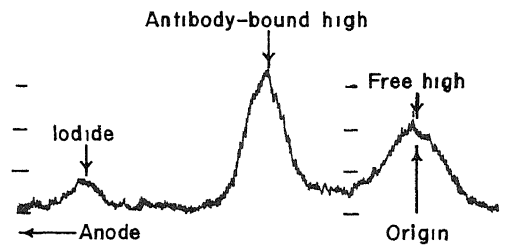


Fig 7 4 Paper radiochromatoelectrophoretogram of mixture of HGH-^{131}I diluted HGH antiserum and human plasma after incubation at 4° C Free HGH-^{131}I is absorbed to the paper at the origin antibody bound hormone has migrated with the plasma protein and iodine-131 released from the labeled hormone during incubation has migrated farther than the plasma proteins (From Yalow and Berson 1968)

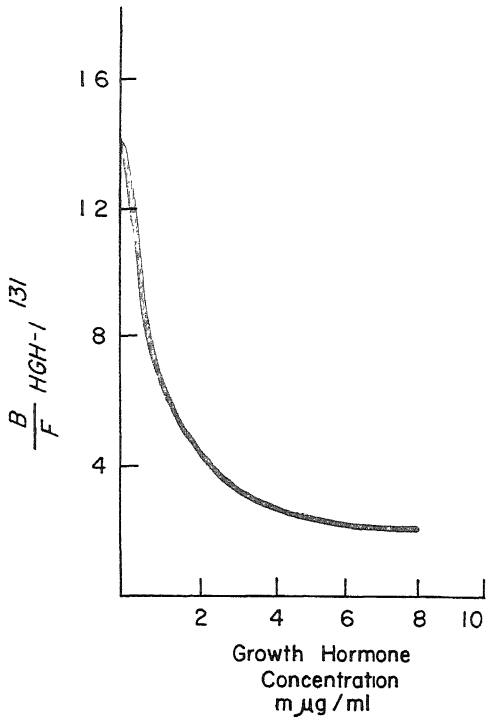


Fig 7 5 Standard curve for HGH assay B/F ratio in standard solutions is plotted as a function of human growth hormone concentration (From Yalow and Berson 1968)

The ratio of bound/free human growth hormone (HGH)-^{131}I is a function of the quantity of unlabeled hormone that is added (Fig 7 5) Once this relationship is established with hormone standards, it can be used to assay plasma levels Radioimmunoassays have been established for insulin, glucagon, FSH, LH, prolactin, TSH, adrenocorticotrophic hormone (ACTH), α-MSH, β-MSH, vasopressin, parathyroid hormone, and calcitonin in addition to growth hormone

Recently, the principle of radioimmunoassay has been applied to the assay of nonprotein hormones This method is based on the presence of specific binding proteins for steroid and thyroid hormones in plasma In normal plasma, a large fraction of these hormones are specifically bound to carrier proteins Competition between labeled and unlabeled hormone for binding has been used and standard curves similar to those of Figure 7 5 have been described Using this method, sensitive assays for thyroid hormone, adrenal steroids, and gonadal steroids have been established

MECHANISM OF HORMONE ACTION

When a hormone reaches its target cell, it must affect the intracellular metabolism in order to modify cell function Peptide and protein hormones are large molecules, however, and will not enter cells readily unless specialized transport systems are present In the case of insulin, the hormone appears to affect cell metabolism by binding to receptors on the surface rather than by entering the cell itself If the hormone is restricted to surface binding, two mechanisms of action suggest themselves, *1)* a hormonally-induced change in membrane permeability for substrates or ions, and *2)* production of a "second messenger" within the cell that will transmit the signal of the hormone (Fig 7 6)

Hormonal control of permeability is well-documented for ions, glucose, and amino acids Increased entry of glucose into cells, which is mediated by a special transport system, is the major way in which insulin controls glucose utilization by muscle Other hormones that change membrane permeability include growth hormone, glucagon, glucocorticoids, estrogen, testosterone, and vasopressin Whether changes in membrane permeability are primary or secondary effects of these hormones is uncertain

Many hormones regulate intracellular metabolism via a second-messenger, 3′-5′-cyclic adenosine monophosphate (cyclic AMP) Sutherland and co-workers (1965) discovered this compound and established its role in hormone action In this system (Fig 7 7), the hormone binds to a receptor on the cell surface that is part of adenyl cyclase, the enzyme involved in formation of cyclic AMP Adenyl cyclase uses ATP as substrate and produces cyclic AMP

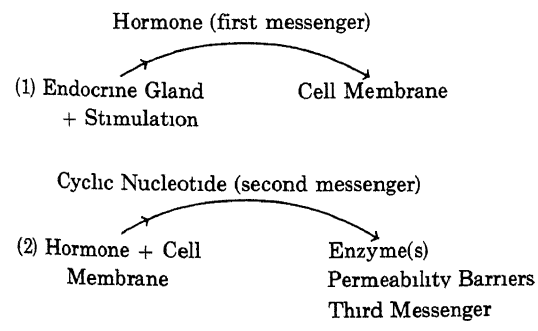


Fig 7 6 The two-messenger system (From Sutherland Oye and Butcher 1965)

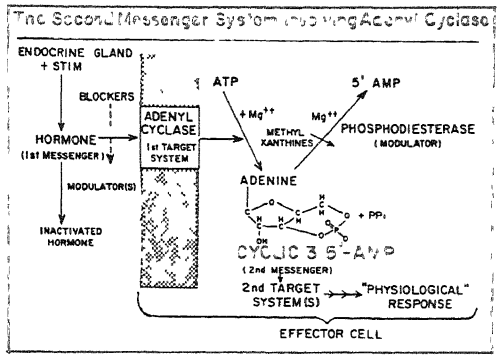


Fig 7 7 The two messenger system involving adenyl cyclase (From Sutherland Oye and Butcher 1965)

and inorganic pyrophosphate as products Intracellular levels of cyclic AMP are most commonly controlled by changes in the activity of adenyl cyclase Destruction of the messenger by phosphodiesterase also effects cyclic levels This enzyme is inhibited by methyl xanthines, such as caffeine, which results in much higher cyclic levels in response to activation of adenyl cyclase Mechanisms involved in the regulation of cell metabolism by cyclic AMP are under intensive study Effects of the second messenger on the following types of events have been described *1)* changes in membrane permeability to ions, water, and amino acids, *2)* release of stored hormone from endocrine glands, *3)* modification in the rate of enzymatic reactions, and *4)* induction of enzyme formation Recently a cyclic AMP-dependent protein kinase has been identified This enzyme may be involved in a number of these effects When protein kinase is activated by cyclic AMP, it catalyzes the phosphorylation of either enzymes or controls transcription of deoxyribonucleic acid (DNA) by histones The wide range of effects which involve cyclic AMP is listed in Table 7 1 As a result of the diverse role of the second messenger, it will be discussed to some extent in every chapter of the endocrine section

Steroid hormones are able to penetrate the cell membrane due to their smaller size and lipid solubility As a result, these compounds are able to modify intracellular metabolism directly A model for the effect of estrogen on the uterine cell is shown in Figure 7 8 The hormone penetrates the cell membrane and binds to a receptor protein in the cytoplasm Estrogen is transported to the nucleus where it

Table 7 1
Extent of regulation by cyclic AMP

Known Effects of Cyclic AMP		
Enzyme or process affected	Tissue	Change in activity or rate
Phosphorylase	Several	Increased
Glycogen synthetase	Several	Decreased
Protein kinase	Several	Increased
Phosphofructokinase	Flatworm	Increased
Tyrosine transaminase induction	Liver	Increased
PEP carboxykinase induction	Liver	Increased
Serine dehydratase induction	Liver	Increased
β-Galactosidase induction	Bacteria	Increased
Overall protein synthesis	Several	Decreased
Glyconeogenesis	Liver	Increased
Ketogenesis	Liver	Increased
Lipolysis	Adipose	Increased
Steroidogenesis	Several	Increased
Water permeability	Epithelial	Increased
Ion permeability	Several	Increased
Calcium resorption	Bone	Increased
Renin production	Kidney	Increased
Contractility	Cardiac muscle	Increased
Tension	Smooth muscle	Decreased
Membrane potential	Smooth muscle	Hyperpolarized
Amino acid uptake	Liver and uterus	Increased
Amino acid uptake	Adipose	Decreased
Clearing factor lipase	Adipose	Decreased
Amylase release	Parotid	Increased
Insulin release	Pancreas	Increased
ACTH release	Anterior pituitary	Increased
TSH release	Anterior pituitary	Increased
GH release	Anterior pituitary	Increased
LH release	Anterior pituitary	Increased
Thyroid hormone release	Thyroid	Increased
Calcitonin release	Parafollicular	Increased
Acetylcholine release	Nervous	Increased
Histamine release	Leukocytes	Decreased
HCl secretion	Gastric mucosa	Increased
Fluid secretion	Insect salivary glands	Increased
Discharge frequency	Cerebellar Purkinje	Decreased
Melanocyte dispersion	Skin	Increased
Aggregation	Cellular slime mold	Increased
Aggregation	Platelets	Decreased
RNA synthesis	Bacteria	Increased
DNA synthesis	Thymocytes	Increased
Cell growth	Tumor cells	Decreased

From Sutherland, 1970

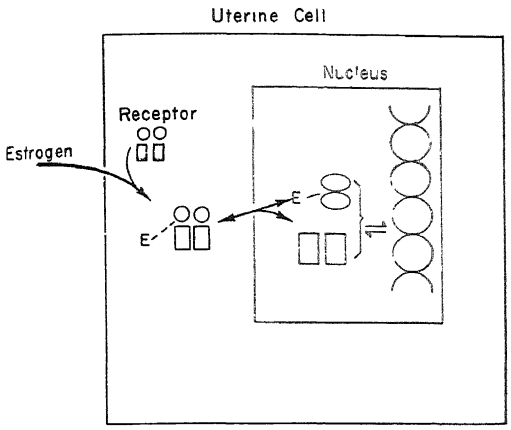


Fig 7 8 Model of estrogen interaction with the uterine cell The steroid binds with a cytoplasmic receptor which transfer it to the nucleus The nuclear receptor facilitates the role of steroids in controlling DNA transcription (Adapted from Gorski et al 1968)

is transferred to a nuclear receptor protein Steroid hormones appear to modify cell metabolism by inducing new enzyme formation This type of control is exerted by regulating the transcription of portions of DNA resulting in formation of messenger ribonucleic acid (RNA) Messenger RNA leaves the nucleus and is translated by the protein synthesis pathway in the cytoplasm Effects of steroid hormone usually require an hour or so to be exerted and may be blocked by inhibitors of RNA and protein synthesis

Steroid hormones are known to exert "permissive effects" on the action of other hormones This term means that the effect of rapidly acting hormones, such as glucagon, depend upon the presence of normal levels of steroid hormones in order to exert their effects Since steroid hormones modify enzyme levels and other cell components, effects of other hormones will change in steroid hormone-deficient animals Control of cell metabolism by rapid acting hormones depends upon the cell constituents that are already present, suggesting that the permissive action of steroids may also depend upon effects on transcription

2 / CONTROL OF THE ANTERIOR LOBE OF THE PITUITARY GLAND

The concept of control of anterior pituitary gland by higher brain centers has crystalized only within the last 10 to 15 years, although elements of the portal vessel-chemotransmitter hypothesis were appreciated as far back as the late 1930's This hypothesis holds that neurosecretory cells in various areas of the hypothalamus secrete chemical substances into hypophysial portal blood vessels whose primary capillary bed lies in the median eminence region These substances called hypothalamic-releasing factors, are carried by the portal vessels and enter pituitary sinusoids where they promote release of hormone from pituitary cells

The anterior pituitary gland probably contains a specific type of hormone-producing cell for each of the following six protein trophic hormones—follicle-stimulating hormone (FSH), luteinizing hormone or interstitial cell-stimulating hormone (LH or ICSH), prolactin (LTH) in some species), adrenocorticotropic hormone (ACTH), thyroid-stimulating hormone (TSH), and somatotrophin or growth hormone (STH or GH) Structures, effects of these hormones on target glands, and general metabolic effects will be dealt with in chapters that follow At the present time, six releasing factors have been identified—FSH-releasing factor (FRF), LH-releasing factor (LRF), corticotropin-releasing factor (CRF), TSH-releasing factor (TRF), GH-releasing factor (GRF), and prolactin-inhibitory factor (PIF) These releasing factors apparently provide a means for a precise set of controls that permit the secretion of trophic hormones in a highly specific and selective way

Some years ago, workers in several laboratories set out to isolate and define the chemical nature of the releasing factors by extracting and purifying minute quantities of biologically active principles from literally millions of fragments of hypothalamic tissue In the last months of 1969, the structure and artificial synthesis of the first releasing factor, TRF, was achieved (Burgus et al , 1970, Folkers et al , 1969) Structures of the remaining factors and their mode of action are under study in a number of laboratories

It is the purpose of this chapter to *a*) briefly review the submicroscopic anatomy of the anterior pituitary, *b*) describe salient features concerning the physiology and chemistry of the releasing factors, *c*) discuss possible modes of action of the hypothalamic factors on the pituitary cell, and finally *d*) indicate the paths that recent studies in this area are taking It is important to note that many comprehensive and excellent reviews on the physiology and chemistry of the releasing factors are available to the interested reader Especially good are Meites (1970), Burgus and Guillemin (1970b), and Schally et al (1968) for the chemistry of releasing factors, Meites (1970), Ganong and Martini (1969), McCann and Porter (1969), and Meites and Nicoll (1966) for various physiological aspects, and Geschwind (1969, 1970) for mechanism of action

Cytophysiology of the Anterior Pituitary

The diversity of cell types in the anterior pituitary gland was first recorded by Schonemann (1892) Since that time results of many hundreds of cytological studies make it clear that the three major groups of cell types, i e , chromophobes, acidophils, and basophils each have subclasses within The chromophobes, which are devoid of any stainable cytoplasmic granules by light microscopy, make up approximately 55% of the cells in the gland The acidophils and basophils constitute the remaining 35% and 10%, respectively These proportions vary depending upon age, sex, and physiological state The acidophils contain cytoplasmic granules which take on acid dyes while the basophils contain granules which stain readily with basic dyes With the excep-

tion of the gonadotropins, the experimental evidence that one cell type produces one particular hormone is convincing

ELECTRON MICROSCOPY

Studies of Smith and Farquhar (1966, 1970) have clearly demonstrated that it is possible to identify at least four of the six cell types primarily on the basis of secretion granule size in their cytoplasms These studies involved examination of subcellular changes in pituitary tissue removed from animals that had been physiologically altered in some predictable fashion For example, thyrotrophs have been identified in normal glands as cells that contain cytoplasmic granules 140 mμm in diameter In glands from thyroidectomized animals, these cells are degranulated and vesiculated Similarly, after castration, cytoplasmic vesicles coalesce to form "signet ring cells" in those cells containing 200-mμ diameter granules These cells are therefore considered gonadotrophs In the lactating female rat, LTH secretion is suppressed if the young are separated from the lactating mother Some 10 to 12 hr after separation there is an accumulation of mature secretion granules (600 to 900 mμm) in cells considered to be responsible for LTH secretion Studies such as these provide not only convincing evidence for the identity of a spe-

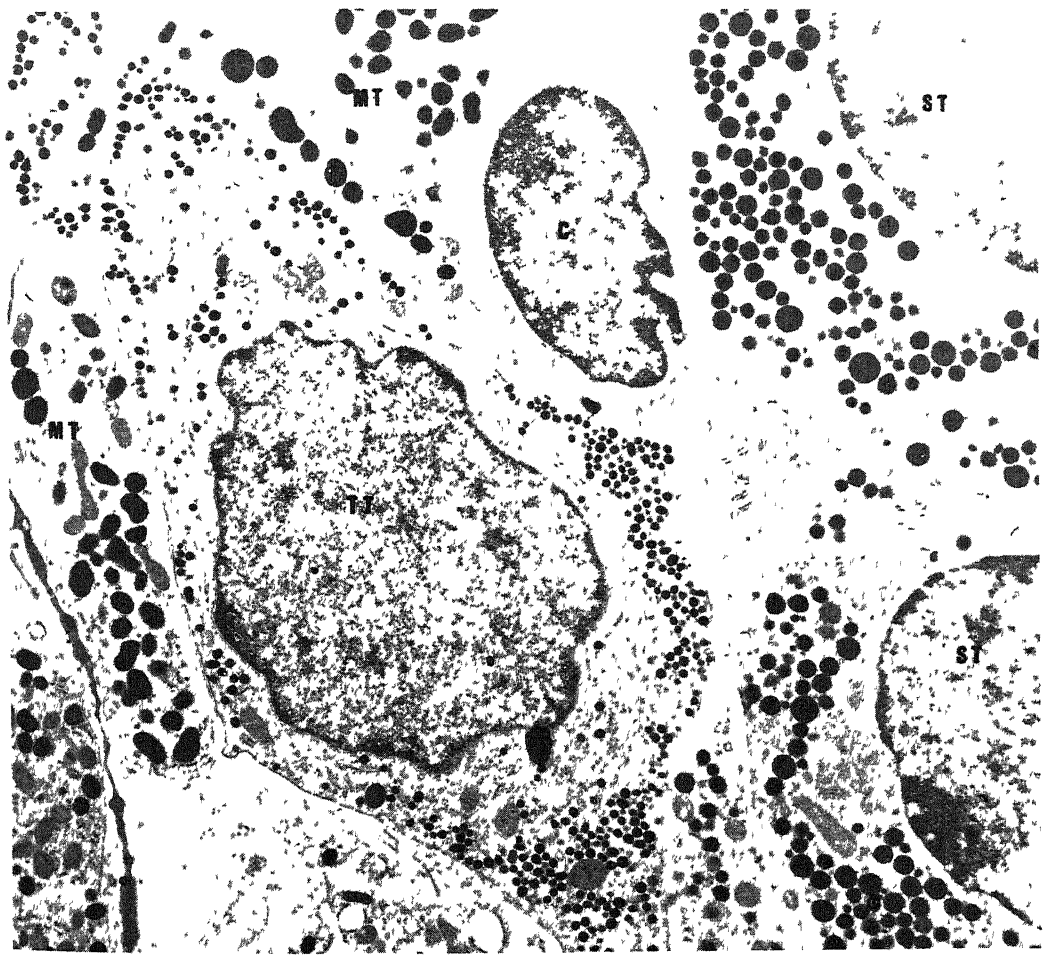


Fig 7 9 Electron micrograph showing a field of cells from the anterior pituitary gland of the rat Several cell types are seen which can be distinguished by differences in the nature of their secretion granules Mammotrophs (*MT*) or prolactin-secreting cells have the largest granules (600 to 900 mμ) and thyrotrophs (*TT*) the smallest (150 mμ) Those of somatotrophs (*ST*) or growth hormone-producing cells are of intermediate size (350 to 400 mμ) A small chromophobe (*C*) or nongranulated cell is also present (× 20 000) (Courtesy of M G Farquhar)

cific hormone-producing cell, but also indicate that hormone synthetic and secretory activity can be correlated with the ultrastructural appearance of the cell (Figs 7 9 to 7 11)

Immunohistochemical techniques have also been used for hormone localization Initially, a fluorescent compound was conjugated to the hormone antibody but more recently an enzyme, horseradish peroxidase, has been used as a conjugate Using this technique Nakane (1970) has found that STH, ACTH, LTH, and TSH were localized in separate cells (see Fig 7 12) FSH and LH, however, were frequently found in the same cell In general, the results from this method confirm that secretion granule size serves as a useful marker for the identification of pituitary hormone-producing cell types

SUBCELLULAR SITES OF HORMONE SYNTHESIS STORAGE AND RELEASE

The rough, endoplasmic reticulum is highly developed in active pituitary cells, suggesting that hormones are synthesized at this site A cell-free pituitary ribosomal system has been found to incorporate labeled amino acids not only into protein, but also into ACTH, STH, and LTH fractions Furthermore, biosynthesis of LTH and STH was found to be most effective in polysomes containing six to seven ribosomal units This finding is in good agreement with the predicted polysome size requirement for the synthesis of proteins with molecular weights of approximately 20 000 (Adiga et al , 1968)

It seems reasonably clear that packaging of

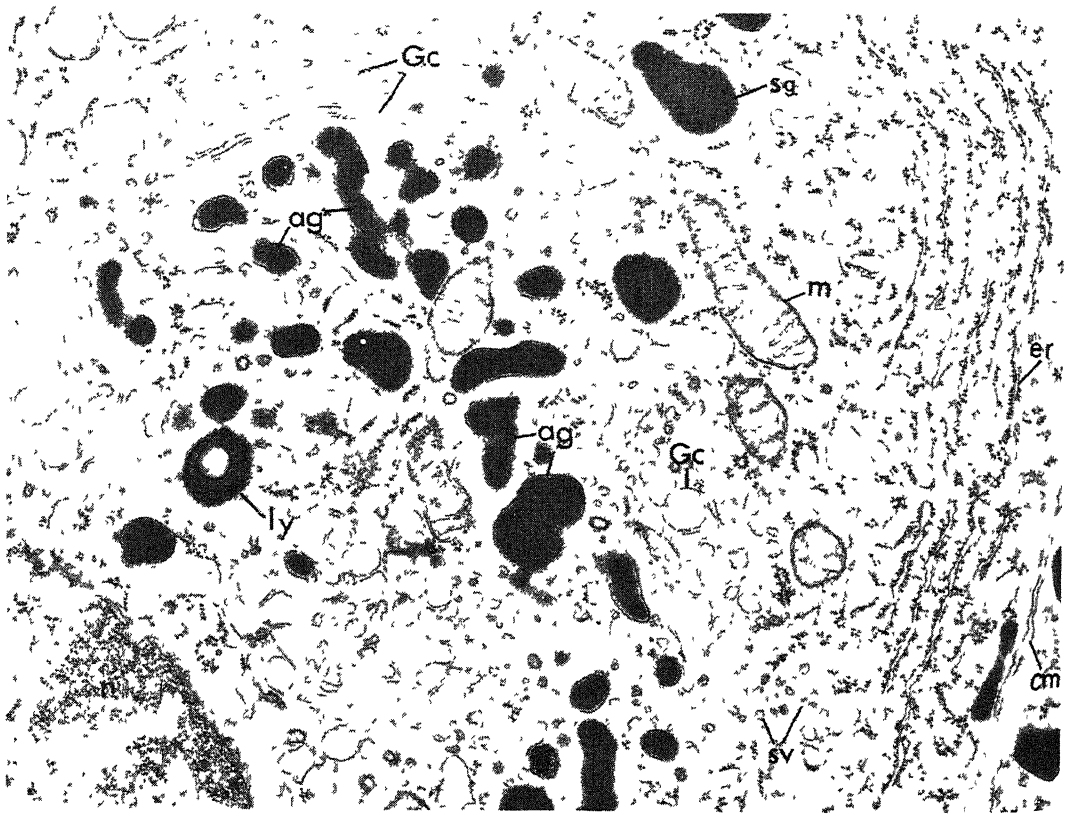


Fig 7 10 Portion of a mammotroph (prolactin secreting cell) from the anterior pituitary of a lactating rat These cells are rapidly synthesizing concentrating and discharging prolactin or lactogenic hormone They show abundant peripheral arrays of rough-surfaced endoplasmic reticulum (*er*) parallel to the cell membrane (*cm*) a large juxtanuclear Golgi complex (*Gc*) and many profiles of forming granules (*ag*) Secretory granules are formed in several steps small (100 to 200 mμ) granules arise within the inner Golgi cisterna pinch off and are found along the concave Golgi face Several small granules merge to form larger aggregates (*ag*) of variable shape and the polymorphous images round up to form ovoid or elliptical mature granules (*sg*) which measure 600 to 900 mμ (× 21 500) (Courtesy of M G Farquhar)

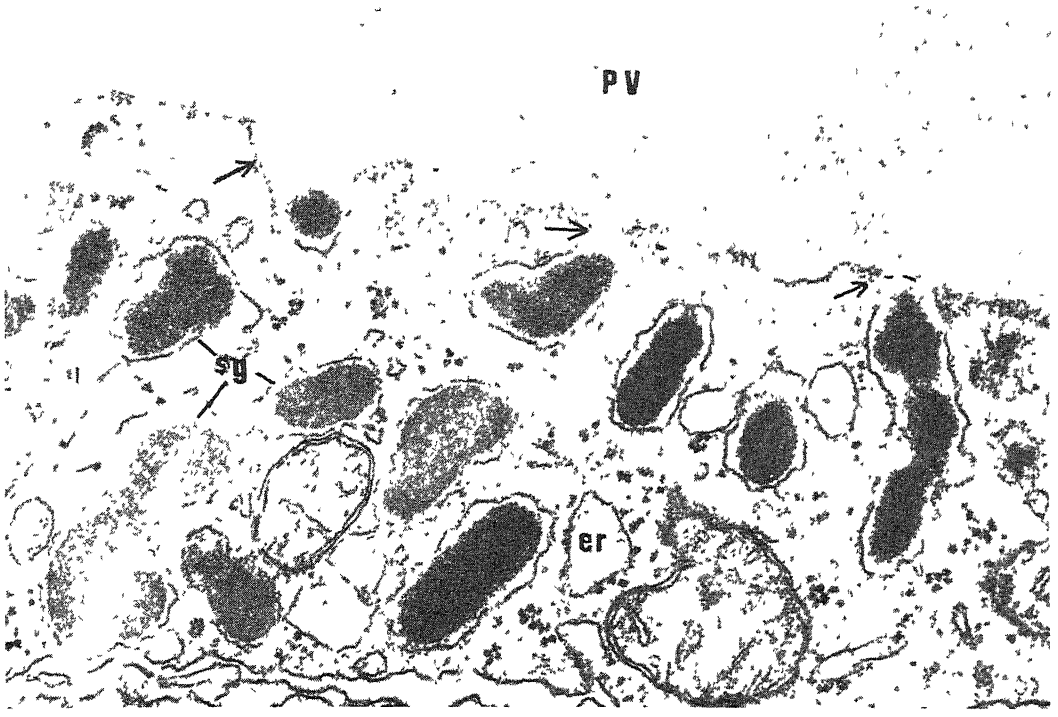


Fig 7 11 Portion of a mammotroph or lactogenic hormone producing cell from a lactating rat A number of secretory granules (*sg*) are lined up facing the perivascular spaces (*PV*) Several are undergoing discharge by exocytosis or fusion of the granule membrane with that of the cell membrane (*arrows*) (× 42 000) (Courtesy of M G Farquhar)

newly synthesized hormone into secretion granules occurs in the Golgi region of the cell Several ultrastructural studies suggest that the development of the Golgi apparatus is directly related to the secretory activity of the cell Since the granule size and form seem to be characteristic for a given cell type, some specificity must exist in the Golgi system of the different cells to determine the arrangement and size of the emerging packet After completion, the granule consists of an electron dense core surrounded by a membrane derived from the Golgi Although there is no direct evidence, it seems likely that some of the carbohydrate moieties of the FSH, LH, and TSH molecules are also attached in the Golgi region Enzymes derived from the Golgi, primarily phosphatases, are also packaged in the secretion granule (Smith and Farquhar, 1970)

Secretion granules of different sizes have been isolated in highly purified form and a major portion of each trophic hormone has been associated with an individual granule fraction (Fig 7 13) There granules, which retain their membrane on isolation, contain biologically active hormone when injected into bioassay animals It is of interest that gonadotropins, associated with granules isolated from basophils, are in a form which makes them extractable in 0 15 M saline Isolated STH granules are stable in saline, however, but unstable in the presence of ethylenediaminetetraacetic acid (EDTA), a chelating agent These results suggest that divalent cations are required for stability of the STH granule Finally, purified granule preparations have been analyzed by polyacrylamide gel electrophoresis Hormones in the isolated granules have mobilities similar to those reported for crude pituitary homogenates or extracts (Hodges and McShan, 1970) Thus, individual hormones are associated with granules of different sizes in a highly specific fashion

On the basis of electron microscopic findings, Farquhar reported some 10 years ago (1961) that hormone discharge from the pituitary cell involved exocytosis or reverse pinocytosis (Figs 7 11, 7 14) This process involves fusion of the granule membrane with the apical cell membrane The electron dense com-



Fig 7 12 Localization of growth hormone in thin section by treatment of section with peroxidase-labeled antibody to growth hormone Reaction products from the substrate 4-Cl-1-naphthol are deposited primarily on secretion granules of growth hormone (*GH*) cells (× 6000) (From Nakane 1970)

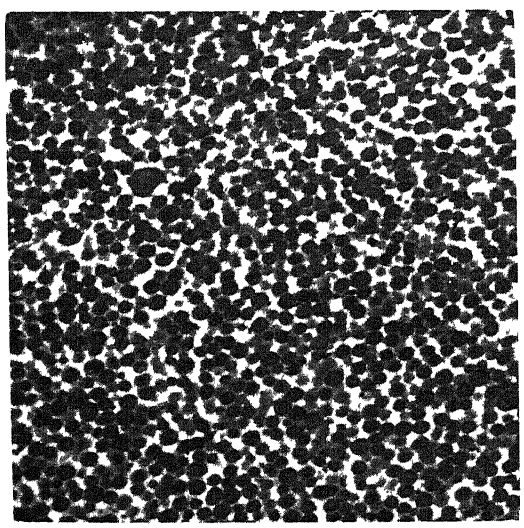


Fig 7 13 Electron micrograph of a thin section through a pellet consisting largely of isolated pituitary secretion granules which are 350 mμ in diamter A majority of the growth hormone activity in the original rat pituitary homogenate was recovered with these isolated granules (× 9000) (From Hymer and McShan 1963)

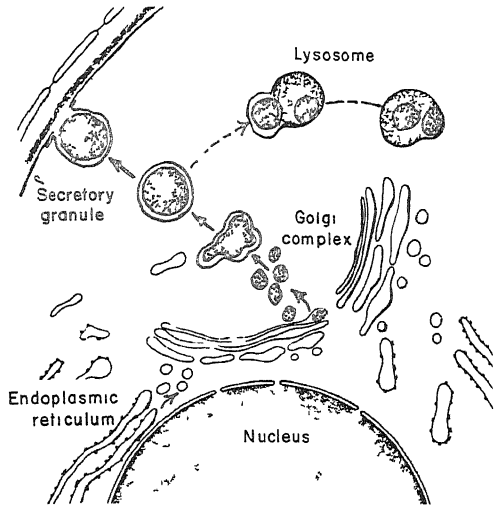


Fig 7 14 Diagram depicting formation of large (600 to 900 mμ diameter) prolactin granules in a pituitary mamotroph by coalescence of smaller granules formed in the Golgi complex During normal lactation the granules are released as shown in Figure 7 11 Upon weaning of the young however demands for LTH secretion are diminished and the granules follow the path indicated by the dotted line Here they are incorporated into autophagic vacuoles and digested by lysosomal enzymes (From Fawcett et al 1969 Redrawn after Smith and Farquhar 1966)

ponent of the granule is released into the pericapillary space Since none of this dense component is found lying free in the perivascular connective tissue or within the vascular endothelium, it has been assumed that the hormone rapidly dissolves or undergoes some physiochemical change prior to transport into the blood A system for degradation of mature hormone is also present in pituitary cells Smith and Farquhar (1966) have shown breakdown of granules containing LTH by lysosomes under conditions where hormone is no longer required

The time required for synthesis and release of hormone molecules has recently been examined Radioautographic studies with the electron microscope indicate that 10 min after administration of radioactive amino acid, label is found in the endoplasmic reticulum Some 30 min later it is concentrated in the Golgi complex, and after 90 min labeled secretion granules are dispersed in the cytoplasm (Racadot et al, 1965) Since the time required for synthesis of a hormone molecule on a polysome is probably no more than 2 min, the above results imply that a relatively long time

must elapse before newly synthesized hormone can be released into the circulation It would be of interest to know if, under conditions of rapid hormone release, intracellular packaging might not be bypassed

Neurchumoral Control

Among the first experiments to provide support for neurohumoral control of anterior pituitary function were those involving *a*) electrical stimulation or lesioning of the hypothalamus, *b*) interruption of the blood vessels between the hypothalamus and the pituitary, or *c*) transplantation of the pituitary gland In 1952, Harris and Jacobsohn transplanted pituitaries into either the median eminence or temporal lobe of hypophysectomized rats Rats with transplants in the median eminence mated and delivered young, but those with transplants under the temporal lobe were anestrus and had atrophic reproductive organs As another example, pituitary cells grown in tissue culture quickly lose their distinctive staining characteristics, but if extracts of hypothalamic tissue are added to the culture vessel, the cells regain normal staining properties (Meites, 1970) Many other experiments could be cited, but all point to the required integrity of the portal blood supply for proper pituitary function (Fig 7 15)

As pointed out by Harris, the region where hypothalamic nerve terminals come into contact with the primary capillaries of the portal vessels is the median eminence Examination of this area with the electron microscope shows a number of granules of varying sizes (1000 to 2100 Å) in addition to the typical synaptic vesicles Preliminary evidence based on gradient centrifugation suggests that these granules may contain releasing factors (Kobayashi and Tokuzo, 1969) Areas of the hypothalamus responsible for synthesis and secretion of individual releasing factors have not been clearly identified Hence, hypothalamic areas, rather than discrete nuclei are discussed when considering control of anterior pituitary function

The following criteria must be satisfied for ultimate proof of the releasing factor hypothesis

1) Secretory elements in the brain must be demonstrated to elaborate a substance within diffusion distances of the primary capillary plexus of the hypophyseal portal vessels

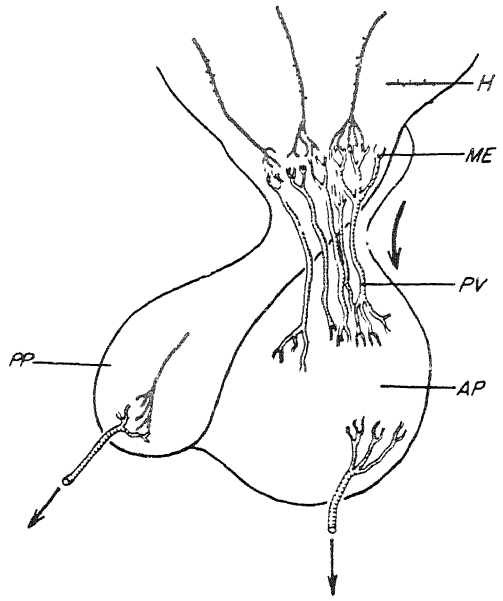


Fig 7 15 Diagram of a midline sagittal section through the base of the hypothalamus and pituitary gland to show the structures concerned in the regulation of anterior pituitary activity Nerve fibers from the hypothalamus (stippled *H*) enter the median eminence (*ME*) of the tuber cinereum to end on the primary plexus of the hypophyseal portal vessels (*PV*) It is postulated that releasing factors pass from these nerve terminals into the portal vessels and are carried into the anterior pituitary (*AP*) to regulate its secretory activity This is probably a similar mechanism to that found in the posterior pituitary gland (*PP*) where peptides are discharged from nerve endings into the general circulation (Diagram and description from Harris 1970)

2) Identification and characterization of this substance in hypophyseal portal blood is required Furthermore, the concentration of the substance should be *a*) greater in portal blood than systemic blood, and *b*) related to varied secretion rates of one of the pituitary hormones

3) When the substance enters the pituitary sinusoids, it causes an alteration in release rate of one tropic hormone

Although only minute amounts of releasing factors would be expected in portal blood, recent experimental evidence suggests that such activity can be detected Porter et al (1970) devised a procedure for collecting stalk blood from the anesthetized rats that involves exposure of the ventral surface of the pituitary via a parapharyngeal approach A dural flap over the gland is cut and the veins draining the pituitary are broken The pituitary stalk is transected and the gland removed The short stump of the pituitary stalk is left protruding

through the diaphragm of the sella, and a polyethylene tube is placed over the stalk in such a manner as to form a tight seal Portal blood can be collected for 4 to 6 hr by applying negative pressure on the cannula In large male rats, 6 μl of blood is collected per minute As indicated above, releasing activity can be demonstrated in this blood

RELEASING FACTOR ACTIVITY IN DIFFERING PHYSIOLOGICAL STATES

Knowledge of the effects of physiological stimuli on secretion rates of releasing factors would aid in an overall appreciation of the regulation of hypothalamohypophysial axis activity This work is difficult because of uncertainties involved in isolation and bioassay of releasing factors and in interpretation of the data In most of these studies, hypothalamic content of releasing factors is measured Tissue content of a hormone represents the balance between rates of synthesis and release It is usually considered that an increase in hypothalamic content of releasing factor represents increased synthesis and secretion if changes of the corresponding hormone in the anterior pituitary and target gland are also observed In those cases where target gland hormones may act directly at the pituitary level, hypothalamic content of releasing factor should remain unchanged Most stimuli, however, appear to act through the hypothalamus and involve changes in secretion rates of releasing factors Experimental data suggest the presence of two feedback mechanisms which modify secretion rates *1*) the "long" loop which involves peripheral target hormones, and *2*) the "short" loop which involves pituitary hormones acting directly on the hypothalamus

Evidence for the long-loop feedback mechanism in the control of FSH and LH secretion stems from the observation that hypothalamic LRF and FRF content is greatly increased after castration This is especially true for the male, but has also been seen to some extent in the female These changes parallel increased synthesis and release of FSH following castration Interestingly, administration of testosterone or estrogen to the castrate male results in a marked depletion of FRF stores Some specificity is indicated by the fact that progesterone administration has no effect on FRF stores in the hypothalamus (Martini et al, 1968) These data argue for the idea that testosterone decreases secretion of pituitary FSH by way of the hypothalamus Results from similar kinds of studies for other hormones have been reported by Meites (1970)

The evidence for the existence of short-loop feedback control has involved systemic administration of pituitary hormones For example, injection of 20 μg FSH/day/5 days to castrated male rats causes a significant drop in pituitary FSH and hypothalamic FRF The specificity of this response is indicated by the finding that injection of LH or ACTH has no such effect In this kind of experiment, the FSH is believed to act directly back on the hypothalamus to prevent secretion of FRF Furthermore, implantation of relatively large doses of pituitary hormones in the median eminence region also results in reduced content of the hypothalamic factor and decreased output of the pituitary hormone Thus, implantation of LH in the hypothalamus decreases LRF and LH, and implantation of ACTH into median eminence significantly depress blood corticosterone levels This method has been criticized because diffusion of hormone from the implantation site undoubtedly occurs, and in fact probably reaches the anterior lobe Electrophysiology has also been used to advantage in some of these problems Work by Beyer and Sawyer (1969) indicates that *1*) "patterns of hypothalamic neural activity can be correlated with pituitary secretion, and *2*) various hormones, including the pituitary trophins, influence the basal level of activity and the responsitivity of hypothalamic neurons " Finally, hypothalamic releasing factors are detectable in the systemic circulation of hypophysectomized rats 1 month after the operation This condition, which is never encountered in the intact animal, is explained by the lack of inhibiting messages from the anterior pituitary Taken together, the results from experiments utilizing these different techniques strongly support the notion that a "short" feedback mechanism functions in the regulation of pituitary secretion It is an attractive mechanism for the control of secretion of those hormones that do not have a peripheral target gland and are therefore not regulated by the more traditional feedback mechanisms It is not clear whether this system operates by the systemic circulation or by transport up the pituitary stalk

In the case of regulation of ACTH and TSH secretion, fairly good evidence can be mustered for an additional feedback control directly at the pituitary level Data on regulation of gonadotropin secretion are controversial, but a basomedial hypothalamic receptor for testosterone inhibitory feedback seems fairly well-established Estrogen may also have a primary action on the basomedial hypothalamus, with additional secondary effects on the anterior pituitary (Davidson, 1969) In the case of growth hormone secretion, a number of studies point to the existence of a specific glucose-sensitive growth hormone regulatory system Thus, hypoglycemia or falling blood glucose will stimulate STH secretion Other stimuli, such as sleep, or those related to protein metabolism (e g , arginine and several other amino acids) will also stimulate STH secretion, but whether these stimuli are specific or merely act as stressors is debatable While it is clearly accepted that STH secretion is under control of the hypothalamus, the relationship of the above stimuli to the hypothalamus remains to be defined (Glick, 1969)

There is a considerable body of evidence to indicate that brain catecholamines participate in the regulation of gonadotropin secretion Fuxe and Hokfelt (1969) have demonstrated, by a histochemical technique, the presence of dopamine-containing neurons whose cell bodies lie in the arcuate nuclei Axons from these neurons project to the external layer of the median eminence in close proximity to the portal blood vessels Interestingly, their dopamine content fluctuates under conditions of altered gonadotropin secretion Dopamine could participate in gonadotropin secretion by *a*) directly stimulating release of pituitary FSH and LH, *b*) potentiating the effect of the gonadotropin releasing factors, or *c*) functioning as a synaptic transmitter to produce the release of FRF or LRF Data from Kamberi et al (1970) indicates the latter possibility is most likely For example, basal release of FSH from pituitaries incubated in flasks containing median eminence fragments was unaltered when serotonin, epinephrine, or norepinephrine were added However, FSH release was significantly stimulated by addition of dopamine Furthermore, addition of dopamine to pituitaries incubated alone failed to alter FSH release In more recent experiments, utilizing an in vivo approach, injection of dopamine into the third ventricle of hypophysectomized rats caused dramatic and specific increases in plasma levels of LRF (Schneider and McCann, 1970) Furthermore injection of estradiol into the third ventricle 2 hr prior to dopamine administration would block this response

Finally, it should be obvious to the reader that the loop systems discussed in this section are extremely complex In future work it will be important not only to identify sites of feedback action and multiple actions of loops, but also the relative importance and interplay of each of the loops in the physiological regulation of anterior pituitary function

ASSAY OF RELEASING FACTORS

An in vitro pituitary incubation system, sometimes called the short-term incubation system, was originated by Saffran and Schally (1955) and was later modified by Guillemin, Meites, and others Basically the method involves incubation of rat anterior pituitary glands in small flasks or beakers placed in a Dubnoff metabolic shaker maintained at 37° C with continuous oxygenation Under these conditions, tissue and cell viability appears relatively normal after 1 or 2 hr incubation It is also possible to maintain explants of pituitary tissue in organ culture Tissues handled in this manner remain functional for a week or longer

The results of studies with the short-term incubation system show that addition of crude acid extracts of hypothalamus causes the release of biologically detectable quantities of all the anterior lobe hormones with the exception of LTH Since LTH levels are high in media taken from glands incubated in the absence of hypothalamic extracts, it was concluded by Meites (1970) that the hypothalamus contained a prolactin-inhibiting factor (PIF) This idea is entirely in keeping with the observations that when pituitaries are transplanted to the kidney capsule, *a*) the predominant surviving cell type is the mammotroph, and *b*) that since the gland has been removed from its hypothalamic control, there is an enhanced release of LTH from the graft

Hypothalamic extract-induced release of hormone from the in vitro pituitary is specific as suggested by the following lines of evidence First, the quantity of released hormone is dependent upon the concentration of extract used, i e , it follows a dose response curve

Second, if crude acid extracts of cerebral cortex tissue are added, the amount of hormone released is often no greater than "basal" or endogenous release levels Furthermore, no PIF is found in cerebral cortical extracts Third, only one hormone is released from the tissue when a single purified factor is added The experiments of Guillemin and Vale (1970) demonstrate this important point If purified TRF is added to the in vitro system, TSH output is greatly increased above endogenous release levels while ACTH and LH are not (Fig 7 16) This specificity is further demonstrated by the findings that *a*) thyroxine inhibits the TRF effect on release, *b*) TSH release is related to dose of TRF, *c*) oxygen is required for the full effect of TRF, and *d*) pituitary response to TRF is decreased at 0° C

Although hormone release is clearly demonstrated in the in vitro procedure, the method provides only a static view of the process A potentially useful method for studying the kinetics of hormone release in vitro has recently been proposed (Serra and Midgley, 1970) In this method a single anterior pituitary gland is incubated in a 350-μl capacity vial constructed in such a way that the gland can be superfused in continuous fashion and the perfusate analyzed for hormone content (Fig 7 17) With this set-up, it has been shown that a single pituitary gland will respond to hypothalamic extract by release of LH, that the response is dose related, that the response consists of a brief release essentially limited to the time of extract administration followed by a fall to a constant basal release rate and finally

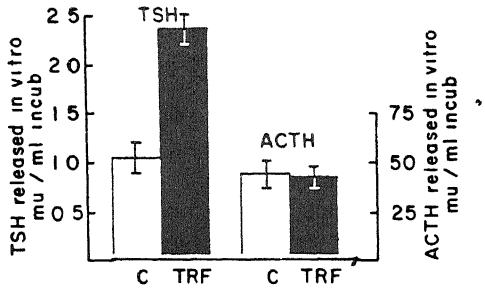


Fig 7 16 Specificity of the response of the rat pituitary incubated in vitro With the use of paired hemipituitaries addition of *TRF* stimulates release of *TSH* whereas no increase in the release of *ACTH* is observed when both *TSH* and *ACTH* are measured in aliquots of the same incubation medium Sensitivity of the assay method for *TSH* 0 02 mU, for *ACTH* 0 1 mU (From Guillemin and Vale 1970)

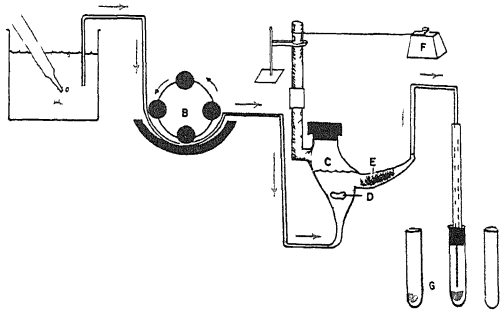


Fig 7 17 Incubation design *A*) buffer container with 95% oxygen and 5% carbon dioxide continuously bubbled *B*) peristaltic pump *C*) incubation chamber containing 200 μl of buffer *D*) intact rat anterior pituitary gland *E*) glass wool filter *F*) mixer producing vibrations of the incubation chamber and *G*) test tubes (From Serra and Midgley 1970)

that the same gland can be repeatedly stimulated to release measurable quantities of hormone (Fig 7 18) Techniques such as these clearly permit new approaches to the study of release kinetics of the pituitary trophic hormones

IN VIVO ASSAY FOR RELEASING FACTORS

A number of different in vivo assay methods have been used to check for the presence of releasing factor activity One type, the "pituitary depletion assay" is based on the general principle that injection of a preparation containing one or more releasing factors brings about a depletion of pituitary hormone content related to the amount of factor injected (Motta et al, 1970) Another general type of assay utilizes the principle that in vivo administration of a releasing factor causes an increase in the blood level of a particular pituitary hormone A modification of this method involves measurement of the change in target gland hormones Each assay method that is used, in vitro as well as in vivo, has advantages as well as disadvantages Lack of precision, due not only to variability in pituitary response to a releasing factor but also to variability in assay of released pituitary hormone, is a common problem

Recently Gay et al (1970) have developed a method for constant monitoring of plasma LH concentrations in individual animals Using a sensitive radioimmunoassay for rat LH, these workers have been able to analyze the dynamics of LH release following one or more

injections of crude rat hypothalamic extract In this method, the castrated female rats are treated for 3 days with estrogen and progesterone Increases in circulating LH in response to injection of hypothalamic extracts occur within the first 2 min after injection and maximal values occur 8 min after extract administration (Fig 7 19) These preparations are capable of responding to repeated injections of hypothalamic material Moreover, it is possible

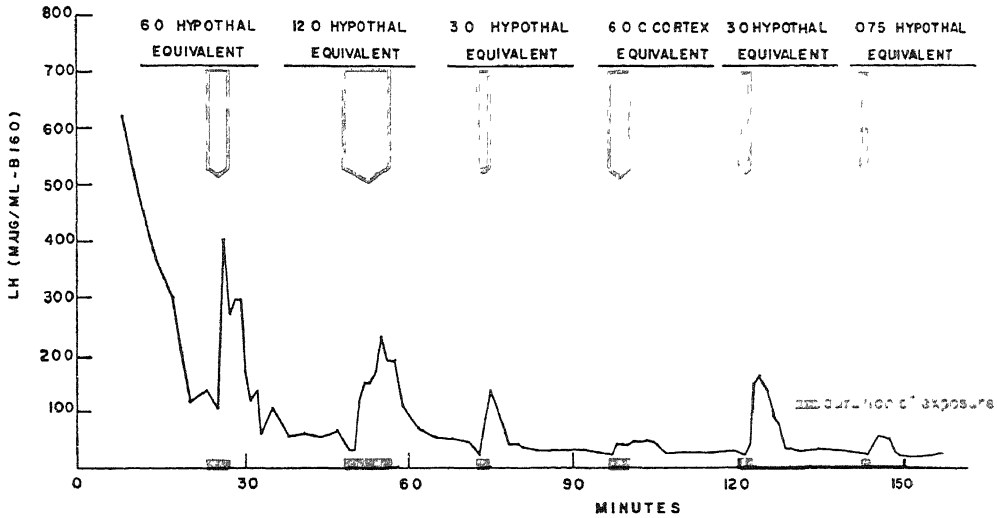


Fig 7 18 Concentration of LH in the superfusate following exposure in vitro to hypothalamic extracts for 4 8 2 and 0 5 min Exposure to 6 cerebral cortex equivalents was done during a 4-min period Samples were collected and measured at 1- or 3-min intervals Superfusion rate using the apparatus shown in Figure 7 17 was constant at 200 μl /min LH was measured by radioimmunoassay in 10 or 20 μl aliquots of superfusate (From Serra and Midgley 1970)

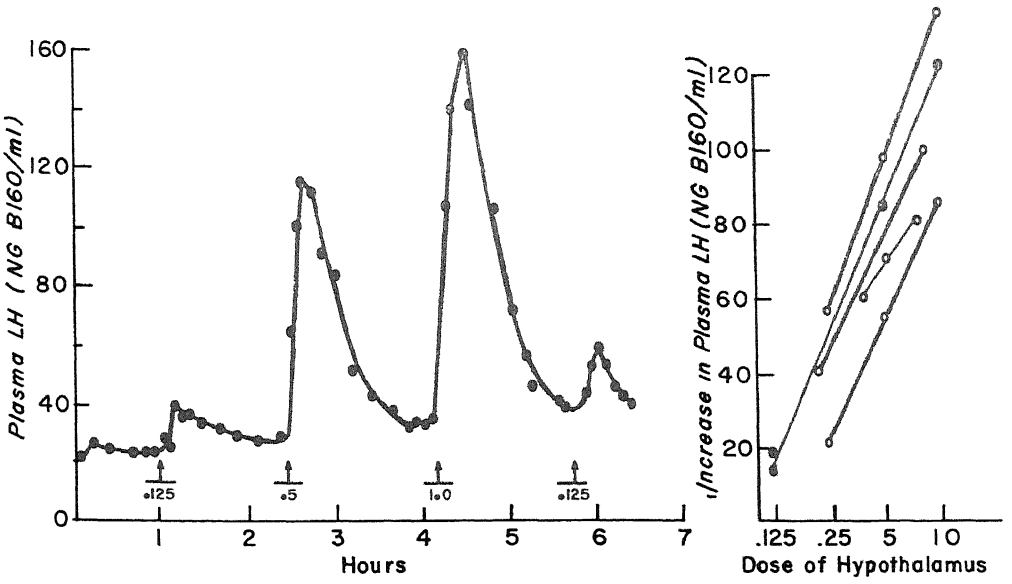


Fig 7 19 Demonstration of the dose related increase in plasma LH following 4 successive injections of 0 125 0 5 1 0 and 0 125 hypothalamic equivalents into an estrogen progesterone-treated castrate rat The observed responses of the animal have been replotted in the right hand portion of the figure to illustrate the linear relationship between the increase in plasma LH and the log-dose of hypothalamic extract injected Responses observed in other similarly prepared rats are indicated by the open circles in the right-hand portion of the figure From data obtained in these experiments it is calculated that LH has a half-life (or disappearance rate) of 23 min (From Gay et al 1970)

to show an increase in plasma LH which is related to dose of extract administered This method, which clearly provides a useful way for evaluating the kinetics of hormone release in vivo, will most certainly be used in future work

CHEMISTRY OF THE RELEASING FACTORS

Purification and chemical isolation of the releasing factors is currently a major goal in a number of different laboratories This approach has recently led to determination of the structure and synthesis of TRF As the initial purification step, it is a common procedure to extract frozen tissue or acetone powders of lyophilized hypothalamic tissue with dilute acid Glacial acetic acid is used most frequently Such extracts are then subjected to preparative gel filtration on columns of Sephadex G-25 The order of elution from the column is GRF, CRF, TRF, FRF, PIF, and LRF Vasopressin emerges last (McCann and Porter, 1969) In some cases there is considerable overlap of activities Further purification is often obtained by chromatography on ion exchange cellulose In the final stages electrophoresis, partition chromatography, or counter current distribution techniques have been used The enormity of the tasks involved can be appreciated when one considers that a typical yield of "highly purified" releaser material is on the order of 1 to 10 mg from a starting material of several hundred thousand hypothalami

The procedures listed above are those commonly used for the isolation of peptides From their behavior, it seems clear that the releasers are low molecular weight peptides, probably on the order of 1000 MW Except for TRF, all of the releasers are inactivated by proteolytic enzymes Most of the partially purified factors are stable when heated to 100° C for a few minutes Highly purified materials are exceedingly potent in terms of biological activity It is not uncommon to find that doses in the nanogram or picogram range will have measurable effects As a result, hypothalamic factors will release many times their own weight of pituitary hormones, i e , they show a so-called "multiplier effect " The [(amount of hormone produced)/(amount of releasing factor added)] is often several thousand, or higher

TRF. The structure of both procine and ovine TRF is (pyro) Glu-His-Pro (NH_2) (Fig 7 20) Pyroglutamic acid, (also known as 2 pyrrolidone-5-carboxylic acid) forms the N terminus while proline amide forms the C terminus portion of the molecule (Burgus et al , 1970a, Folkers et al , 1969) Synthetic and natural TRF have essentially identical biological activities Both the in vivo and in vitro responses to synthetic TRF are inhibited by administration of triiodothyronine (Bowers et al , 1970a) Finally, in a number of recent studies (Bowers et al , 1970b, Martin and Reichlin, 1970), a dramatic rise in plasma TSH was seen 2 to 5 min after intravenous injection of the synthetic molecule into mice or rats (Fig 7 21)

Minor modifications in the chemical structure of TRF decrease or eliminate biological activity For example, simple tripeptides of glutamic acid, histidine, and proline in all

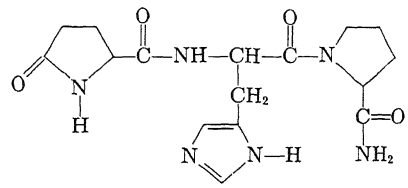


Fig 7 20 Structure of TRF (From Burgus et al 1970)

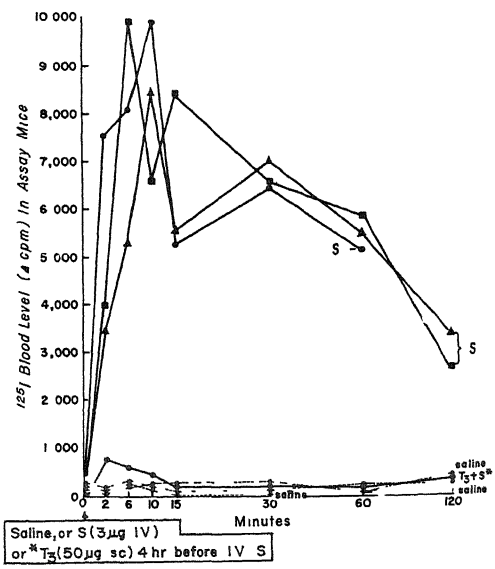


Fig 7 21 Effect of (pyro) Glu His Pro (NH_2) (S) on plasma TSH levels of adult male rats Plasma TSH levels determined by bioassay in mice and recorded as ^{125}I blood level (c p m) Each result is the mean obtained from 3 to 4 mice ± S E Note how prior administration of T_3 blocks the response to (S) (From Bowers et al 1970a)

their possible sequences are not biologically active The addition of a single CH_3 group on the amide of proline reduces the activity to approximately 1/8 that of the synthetic compound The synthetic tripeptide (pyro) Glu-His-Asp (NH_2) is not active, nor is (pyro) Glu-His (NH_2) or (pyro) Glu-Pro (NH_2) All of these findings (Bowers et al , 1970a) clearly show that the entire structure is required for full biological activity The importance of the artificial synthesis of TRF cannot be overemphasized The following is taken from a recent paper (Bowers et al , 1970b) "The simplicity of the structure engenders delight as well as amazement It is noteworthy that this is the first known chemical structure of one of the hypothalamic hormones which have been sought for so long Studies of procine and ovine TRH are equally exciting since the results have been obtained independently and confirm one another and also because they represent once more unequivocal evidence that the hypothalamus secretes specific neurohormonal substances which control the release of anterior pituitary hormones It is hoped that the structure of TRF is a key which will help reveal the structure of the other hypothalamic releasing hormones "

GRF A peptide with GRF activity has been purified and appears to be homogeneous by thin layer chromatography and electrophoresis (Schally et al , 1970a) Ninety-eight per cent of its dry weight could be accounted for by the amino acids shown in Table 7 2 GRF is an acidic polypeptide with a molecular weight of about 1600, that is split by trypsin, chymotrypsin, or papain into small fragments which are not biologically active Schally reports that as little as 0 1 picogram of this material will cause release of growth hormone from pituitaries incubated in vitro Purified GRF is active in vivo as determined by the pituitary depletion assay method The active dose administered to the assay animal was 1 ng

During the course of screening fractions from Sephadex columns for GRF activity by in vitro assay, fractions adjacent to the GRF peak which would diminish output of bioassayable growth hormone by approximately 50% were found (McCann and Porter, 1969) This material was called growth hormone inhibiting factor (GIF) A dual control for growth hormone release by GRF and GIF can thus be easily envisaged Whether release of all pituitary hormones are under such dual control is unknown at the present time

Table 7 2
Amino acid composition of porcine GRF

Amino Acid[a]	Molar Ratio[b]	Probable Residues[c]
Lysine	1 0	1
Histidine	0 9	1
Ammonia	2 1	2
Serine	1 4	1
Glutamic acid	4 9	5
Alanine	3 8	4
Valine	1 1	1
Leucine	2 0	2

From Schally et al , 1970
[a] Amino acids account for 98% dry weight
[b] Accepting lysine as 1 0
[c] To the nearest integer, assuming 15 amino acid residues

Gonadotropin Releasing Factor. There is little doubt that there are substances in extracts of median eminence tissue which will induce release of LH and FSH from pituitaries in vitro or raise plasma levels of these hormones in vivo Until recently it was generally held that release of these gonadotropic hormones was controlled by discrete molecules, viz FRF and LRF, which could be chemically separated (Fig 7 1) However, Schally et al (1971) has recently suggested that a single hypothalamic factor controls the secretion of both FSH and LH His most highly purified preparation has both LRF and FRF activity, yet only one amino acid sequence (pyro) Glu-His-Trp-Ser-Tyr-Gly-Leu-Arg-Pro-Gly-NH_2 The structure was confirmed by chemical synthesis The synthetic decapeptide apparently shows the same kinds of responses in biological activities as the natural material Thus it *a*) raises the concentration of LH in the plasma of ovariectomized rats pretreated with estrogen and progesterone, *b*) raises plasma levels of FSH and LH in castrated male rats pretreated with testosterone propionate, and *c*) stimulates release of FSH and LH from the in vitro pituitary

PIF. Certain fractions obtained from Sephadex and ion exchange columns contain prolactin-inhibitory-factor activity The most commonly used assay procedures have been either the inhibition of prolactin release from the in vitro pituitary or prevention of suckling-

induced depletion of pituitary prolactin Essentially nothing is known about the chemistry of PIF Data from the laboratories of MacLoed and Daughaday (In Meites, 1970) indicate that epinephrine and norepinephrine may inhibit prolactin secretion in vitro, however, as pointed out by Meites (1970) there are many reasons to believe that catecholamines themselves are not PIF

CRF CRF has not been prepared in pure form Two difficulties have impeded this work *1)* CRF activity appears to decrease, rather than increase, in the later stages of purification, and 2) bioassays are difficult since the stress of the assay procedure interferes with the determination of CRF activity (Vernikos-Danellis and Marks, 1970) In fact, McCann and Porter (1969) claim that much of the confusion concerning regulation of ACTH release can be attributed to the indirect methods used in evaluating release rate

In spite of these reservations, there appears to be little question that fractions containing CRF activity can be isolated from median eminence tissue Most of the evidence indicates that the active molecule is a small polypeptide Data from a number of different kinds of experiments also show that this material is not vasopressin

Experiments performed by Saffran et al (1955) clearly indicated that there were substances in neurohypophyseal extracts, distinct from vasopressin, which also stimulated release of ACTH from the anterior pituitary These workers found activity in two types of CRF preparations from porcine posterior pituitary powders and they were designated α and β CRF β CRF contains serine and histidine, in addition to the amino acids of lysine vasopressin It is not known at this time whether α or β CRF constitutes a "true physiological" corticotropin-releasing hormone

HYPOTHALAMIC RELEASING FACTOR ACTIVITY IN MAN

Releasing factor activity does not appear to be species specific, since highly purified ovine and porcine-releasing factors are active in humans Moreover, active extracts have been prepared from human hypothalami. As examples, a single injection of FRF into amenorrheic or anovulatory women induced a rise in plasma FSH within 30 min after administration; plasma LH levels rise rapidly following infusion of LRF in children or adult patients, small amounts of procine TRF administered to human athyreotic cretins cause a significant elevation in plasma TSH 3 min later and circulating and urinary 17-hydroxycorticoids are raised following intravenous perfusion of hog and beef CRF extracts in the human (Harris, 1970, for detailed references)

Two recent abstracts from Schally's laboratory (Schally et al, 1970a, Bowers et al, 1970b) provide more information on this subject In one study, extracts from 339 human hypothalami were separated on Sephadex G-25 GRF was further purified by free-flow electrophoresis and chromatography on carboxymethylcellulose (CMC) TRF, LRF, and FRF were concentrated by phenol and then chromatographed on CMC The results prompted Schally to claim "It is concluded that in man the release of pituitary hormones is regulated by hypothalamic hormones and that the chemical characteristics (molecular weights, isoelectric points, etc , and in the case of TRF probably the structure) of human hypothalamic hormones are very similar to those of domestic animals "

In the other study (Bowers et al, 1970b), the effect of synthetic TRF was evaluated in 12 normal males One to 800 μg were injected intravenously, blood collected, and levels of TSH, FSH, LH, and STH determined by radioimmunoassay The results showed that *a)* the response was specific in that only serum TSH was elevated, *b)* TSH levels were significantly raised 10 min after injection and reached a maximum at 30 min , and finally *c)* pretreatment with 100 μg T_3 5 to 7 hr before injection of TRF partially or completely inhibited the response to TRF

Taken together, the results clearly point to a new field in which studies with purified and synthetic releases in humans could be usefully applied toward correcting a number of endocrine disorders in man

Mechanism of Action

The study of the mechanism of action of releasing factors on the synthesis and release of hormone by the pituitary cell was a logical and inevitable outgrowth of the physiological and chemical studies described earlier in this chapter In spite of obvious difficulties (e g , availability of pure or purified factors and the

cellular heterogenity of the anterior pituitary), a few consistent findings emerge

RELEASE

Since injection of purified releasing material results in a rapid release of hormone into the circulation, it is generally accepted that a primary action of the releasing factor is the stimulation of hormone release Secretion of catecholamine from the adrenal medulla under the control of acetylcholine or high (K^+) has served as a model for experiments on the anterior pituitary (Douglas, 1968) In the adrenal medulla *a*) Ca^{++} was required for catecholamine release, *b*) acetylcholine stimulated uptake of ^{45}Ca, and *c*) Mg^{++} inhibited the stimulatory response Douglas proposed that catecholamine release was initiated by a decrease in the transmembrane potential of the cell that led to increased membrane permeability allowing a net flux of Ca^{++} into the cells An increase in intracellular Ca^{++} would then somehow promote hormone release Because of the analogy to excitation-contraction in muscle, the entire process was termed "stimulus-secretion coupling" Douglas found that membrane depolarization could be uncoupled from secretion, i e , membrane depolarization could still be induced by high [K^+] under conditions where hormone release was blocked, e g , in a Ca^+-free medium This suggested that the entry of Ca^{++} was a crucial event in the secretion process

When similar experiments were done on anterior pituitaries in vitro, high [K^+] in the incubation medium was found to stimulate release of all the pituitary hormones Calcium was required for "normal" as well as high [K^+]-induced hormone release Measurements of pituitary cell transmembrane potentials from pituitaries incubating in high [K^+], as well as crude hypothalamic extracts, have recently been made As predicted, treatment with high [K^+] produces a reversed, positive transmembrane potential in more than half the cells These positive potentials are explained by a preferential permeability to calcium ions triggered by an initial depolarization of the cell membrane to high [K^+] (Milligan and Kraicer, 1970) It would, therefore, appear that the "stimulus-secretion" coupling model can be usefully applied to problems of hormone release from the anterior pituitary As suggested by Geschwind (1970), the role of Ca^{++} in hormone release may involve *1*) conformational changes in the hormones in the granule and/or plasma membrane, resulting in a change in membrane permeability, *2*) interaction between the membrane of the secretion granule and the inner surface of the cell membrane, *3*) changes in the colloidal state of the cell cytostructure so as to induce granule movement toward the plasma membrane, and *4*) stimulation of contraction of a system composed of microtubules and filaments of actomyosin-like material that links secretion granules to the plasma membrane Kraicer has suggested the term 'excitation-contraction coupling" for this situation

Cyclic adenosine monophosphate (AMP) has been established as a second messenger which mediates many of the actions of a variety of different hormones The following experiments suggest that cyclic AMP may be involved in releasing factor-induced pituitary hormone secretion *1*) Addition of crude stalk median eminence extracts to pituitaries in vitro significantly increased pituitary cyclic AMP levels Specificity in this response was apparent since cerebral cortical extract or lysine-8-vasopressin were without effect *2*) Blockage of cyclic AMP degradation by theophylline augmented hormone release *3*) Purified releasers increased pituitary cyclic AMP levels *4*) Hormone release followed addition of cyclic AMP or its *O'*,*N*-dibutyryl derivative to isolated pituitaries *5*) Calcium ion, although required for release, was not required for releasing factor-induced stimulation of pituitary cyclic AMP

Rasmussen and Tenenhouse (1968) have pointed out that the secretion of a number of stored peptides is mediated by cyclic AMP and that Ca^{++} is required for the action of this nucleotide They suggest that the conversion of a small pool of adenosine triphosphate (ATP) to cyclic AMP decreases calcium chelation, making more calcium available for stimulation of peptide secretion These observations have prompted Geschwind to propose the following useful scheme for releasing factor action "an adenophypophysiotropic factor activates adenyl cyclase and/or modifies the cell membrane to allow Ca^{++} entry If the only effect is on adenyl cyclase, then either cyclic AMP modifies the membrane to allow Ca^{++} entry

or membrane Ca^{++}, by being freed from a Ca^{++}-ATP complex and entering the cytosol, passively allows the entry of extracellular Ca^{++} Cyclic AMP then acts, in some still un-

known fashion, to promote release, and for this effect both Ca^{++} and energy are required Cyclic AMP is hydrolyzed by an intracellular cyclic nucleotide phosphodiesterase whose synthesis may be induced in specific cells by the appropriate target organ hormones (e g , dexamethasone and thyroxine) "

Finally, addition of target hormone to pituitaries in vitro has been shown to inhibit high $[K^+]$-induced hormone release This has been shown for thyroxine (Vale and Guillemin, 1967) and corticosterone (Kraicer et al, 1969) on release of TSH and ACTH, respectively Furthermore, thyroxine will inhibit TRF-induced release of TSH, and incubation of rat adenohypophysis in the presence of physiological concentrations of corticosterone (0 25 to 50 μg /ml) suppress hypothalamic extract-induced release of ACTH (Kraicer and Milligan, 1970) Such experiments indicate that these hormones may exert additional feedback control directly on the pituitary cell membrane

SYNTHESIS

In general, two experimental approaches have been used to examine the possibility that releasing factors act within the pituitary cell by controlling the rate of hormone synthesis In one approach, the bioassayable hormone content of the pituitary is measured after in vivo or in vitro administration of a releasing factor to determine if *a*) the amount released, *b*) the pituitary content itself, or *c*) the sum of the two exceed the amount originally present in the gland Evidence from experiments of this kind indicate that hypothalamic factors stimulate hormone synthesis For example, *1*) extracts of median eminence stimulate release of TSH and ACTH from pituitary autografts implanted beneath the kidney capsule 67 days previously although the gland presumably contains little or no TSH or ACTH, *2*) intracarotid injection of rat median eminence extract produces a 5-fold increase in pituitary ACTH content 2¼ min later, *3*) pituitaries taken from ovariectomized rats pretreated with estradiol and progesterone contain 40% more FSH (in the gland + medium) after treatment with RFR than that recovered in control incubations of 4-hr duration, and *4*) incubation of pituitaries from estrogen-progesterone-blocked ovariectomized female rats for 2 hr in LRF caused a maximum release of approximately 5 μg /mg pituitary, almost twice that originally found in the glands

The results of experiments involving changes in bioassayable content are subject to criticism because of the possibility that one is measuring the activation of a precursor hormone molecule rather than its de novo synthesis As an example, peptide bond cleavage occurs in the conversion of proinsulin to insulin Since a number of specific proteases are encountered in pituitary tissue, and since their levels fluctuate depending upon the physiological state, a similar conversion could occur under the influence of a releasing factor Moreover, in the case of the glycoprotein hormones, one could also imagine a control step in which attachment of carbohydrate to inactive polypeptide would be sufficient for activation (Geschwind, 1969, 1970)

The other approach to studying the effects of releasing factors on hormone synthesis has been to measure the incorporation of radioactive amino acid into a pituitary hormone The majority of such experiments have yielded negative results The careful study of Salmi and Geschwind (1967) is representative Pituitaries from ovariectomized rats pretreated with estrogen and progesterone were incubated for 4 hr in the presence of hypothalamic extract and L-^{14}C-leucine or D-^{14}C-glucosamine Although the extract promoted definite release of LH, the incorporation rate of both labeled precursors into LH was not different from that found in control incubations In one of the few positive experiments of this type, Wilber (1970) has reported that addition of synthetic TRF to rat pituitaries in vitro results in a 4-fold increase in incorporation rate of ^{14}C-glucosamine into pituitary plus medium TSH At the present time, the question of whether a releasing factor may increase hormone synthesis as a primary or secondary event remains difficult to answer Many workers believe that synthesis occurs only after substantial hormone release suggesting that synthesis of new hormone is triggered by simple mass action

Future Studies

What are the directions that future studies in this area will take? First, it must be obvious that improved procedures for the bioassay of the releasing factors will be crucial for advancement In this regard, the use of quantitative and sensitive radioimmunoassays are be-

coming more routine Secondly, it is highly desirable that the activity of a releasing factor preparation be defined in some sort of standard unit Not only would this yield more comparable results in individual laboratories, but would also be extremely useful in the quantitation of hypothalamic contents of releasing factors in a variety of physiological states Thirdly, work on the purification and isolation of the releasing factors will continue to dominate the field The successful synthesis of TRF has been a major accomplishment to date Announcements of the structures of other factors should be forthcoming Burgus and Guillemin (1970b) have, in fact, observed "the hypothalamic-releasing factors may all be members of a closely related series of small polypeptides, possibly with a protected N terminus and an amidated C terminus " Some initial work on the effects of these factors in humans has been done, and the possibilities for extensive clinical investigations look promising indeed Tests of pituitary function, induction of ovulation in certain patients or induction of growth in selected individuals with growth hormone deficiency may be possible with improved knowledge of releasing factor chemistry Antagonists of releasing factors can be synthesized, and such diseases as diabetes or carcinoma of the breast may be treatable by a "medical" hypophysectomy If antagonists to the gonadotropin releasing factors could be synthesized it should, of course, be also possible to prevent conception Finally, the mechanism of action of these factors on the pituitary cell will continue to be an important area of investigation In this regard, attempts are being made in the authors' laboratory to isolate relatively homogenous populations of different types of pituitary cells An example of fractions containing primarily small chromophobes, and acidophils is shown in Figure 7 22 *a*, *b*, respectively Basically the method used involves sedimentation of a trypsinized cell suspension through a shallow (1 to 3%) gradient of bovine serum albumin maintained at unit gravity Under these conditions, the sedimentation rate depends largely on the size of the cell Such isolated cells are viable, they synthesize protein and ribonucleic acid (RNA) and contain detectable quantities of hormone It will be of interest to determine the physiological and biochemical effects of purified releasing factors on each preparations

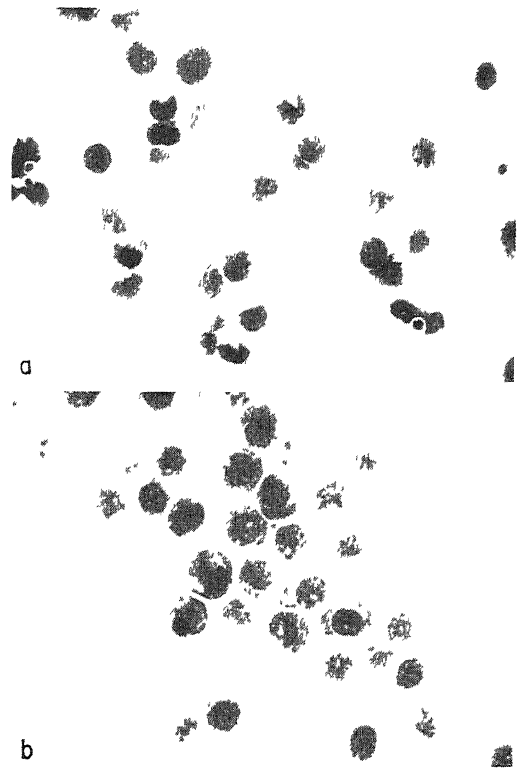


Fig 7 22 Photomicrographs of pituitary cells separated by sedimentation through a shallow protein gradient (1 to 3% bovine serum albumin) maintained at unit gravity according to the method described by Peterson and Evans (1967) Cells in (*a*) are primarily small chromophobes while those in (*b*) are primarily acidophils (× 900)

Acknowledgment In this review, the research from the author's laboratory was supported by the Atomic Energy Commission Grant AT-(30-1)-3827 and a Research Career Development Award from the National Institutes of Health

3 / CONTROL OF THE POSTERIOR LOBE OF THE PITUITARY GLAND

The neurohypophysis is responsible for production and release of two hormones, vasopressin and oxytocin Vasopressin, sometimes called antidiuretic hormone or ADH, is involved in control of water excretion by the kidney and in control of blood pressure Oxytocin stimulates contraction of smooth muscle in the uterus and in the ducts of the mammary gland In this chapter, control of the neurohypophysis and the effects of its secretory products will be described

ORIGIN AND STRUCTURE

The pituitary has a dual origin in the embryo—from the ectoderm in the roof of the primitive mouth as well as from the base of the brain In the early embryo the ectoderm pouch, Rathke's pouch, extends to meet a corresponding hollow diverticulum prolonged downward from the floor of the third ventricle The original cavity of the neural lobe of the pituitary disappears completely in man, so that the pituitary stalk becomes solid The part of the pituitary which originates from Rathke's pouch has a glandular structure and is called the adenophyophysis (anterior pituitary gland) The portion developed from the brain is known as the neurohypophysis (posterior pituitary gland) Each division is subdivided into three parts (Fig 7 23) Thus, the neurohypophysis consists of the expanded distal portion, the infundibular process or neural lobe, the nervous part of the stalk known as the infundibular stem or neural stalk, and the expanded upper end of the stalk, which is the median eminence of the tuber cinereum It is important to realize that all three parts of the neurohypophysis may liberate hormone into the blood stream The adenohypophysis consists of the pars tuberalis, the pars intermedia, and the pars distalis or anterior lobe The neural stalk, together with its sheath, the pars tuberalis, is known as the hypophysial or pituitary stalk The major innervation of the neurohypophysis is derived from the supraoptic and paraventricular nuclei In the infundibular process the fibers of the hypophysial tract branch and interweave and end in close association with capillaries These nerve terminals are loaded with aggregations of a cystine-rich neurosecretory material

A neurosecretory role for nerve cells in the hypothalamoneurohypophysial system was established by the studies of Bargmann (1949, 1954) These cells synthesize vasopressin and oxytocin, transport these substances down their fibers, store them in nerve endings of the infundibular process, and release them into the blood (Fig 7 24) The criteria on which a neurosecretory role is assigned are as follows *1)* identifiable neuronal inclusions share the same distribution as biological activity of the hormones, *2)* interruption of the secretory pathway results in accumulations of secretory material on the hypothalamic side of the lesion, and *3)* appropriate stimuli cause changes in both the number of granules and the biological activity of stored hormone Additional support for synthesis of hormone in the cell body followed by transport down the axon is provided by autoradiographic studies of incorporation of radioactive methionine and cysteine into these nerve cells After injection of the radioisotopes, the cell bodies label rapidly, but the infundibular process does not Ten hours after injection of radioactive cysteine but not methionine, the infundibular process contains large amounts of radioactivity Since vasopressin is rich in cysteine, but contains no methionine, these studies support the concept of synthesis in the cell body followed by axonal transport to nerve endings Neurosecretory material in the storage granules contains hormone bound to a carrier protein Release of hormone from the infundibular process is trig-

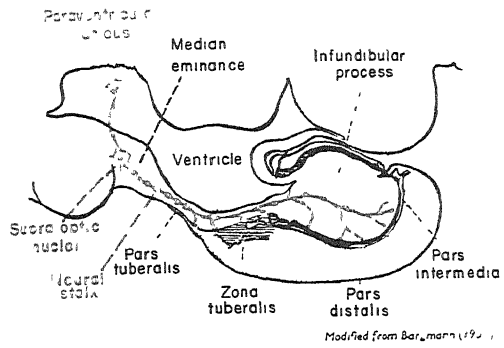


Fig 7 23 Diagram of sagittal section of pituitary and hypothalamus of cat (From Sloper 1966)

gered by an action potential that is transmitted from the cell body As a result, electrical stimulation of the severed stalk of the isolated pituitary can cause secretion (Douglas and Poisner, 1964) After stalk section, about 13% of the total neurohypophysis remains intact and innervated and can release hormone into the blood at a slow rate

HORMONES OF THE NEUROHYPOPHYSIAL SYSTEM

Vasopressin and oxytocin are found in the infundibular process and in the hypothalamus When extracts of the hypothalamus were reported to possess antidiuretic, pressor and oxytocic action (Abel, 1924, Trendelenburg, 1928, and Sato, 1928), considerable skepticism was expressed This reaction was due to the fact that the neural lobe was not recognized as a hormone-storing rather than a hormone-producing gland However, hypothalamic extracts contain only 0 5 to 2 0% of the activity of the neural lobe Analysis of the supraoptic and paraventricular nuclei indicate that the supraoptic nucleus is responsible for production of vasopressin while the paraventricular nucleus accounts for synthesis of oxytocin Synthesis of these hormones in separate nuclei helps to account for differential release under certain conditions, such as suckling and dehydration

Neurohypophysial hormones are of similar structure, consisting of a pentapeptide ring closed by an S-S bridge and a tripeptide side chain (Table 7 3) Almost every vertebrate group has a characteristic hormone or combination of hormones Two hormones are usually found, one resembling oxytocin with a neutral amino acid in the eighth position and another resembling vasopressin with a basic amino acid in this position The hormone combination found in man is oxytocin and arginine vasopressin Vasotocin having the ring of oxytocin and the side-chain of vasopressin produces both antidiuresis and milk ejection when injected into mammals

CONTROL OF VASOPRESSIN SECRETION

Preparation of highly purified and synthetic oxytocin and vasopressin has made possible the analysis of the physiological activities of the whole posterior pituitary extract in terms of these two compounds It seems clear that oxytocin is responsible for the action of pituitary extracts on uterine muscle and on the mammary gland, while vasopressin acts on the blood pressure and water excretion by the kidney Vasopressin has some slight oxytocic activity and oxytocin has slight pressor and antidiuretic action Pressor effects of vasopressin occur only when large amounts of the hormone are administered

Levels of vasopressin in plasma and other body fluids must still be measured by bioassay techniques Antibodies against vasopressin have been produced and a radioimmunoassay procedure has been reported (Permutt et al, 1966) However, this assay is not sensitive enough for measurements of plasma levels Prior to bioassay, hormone must be extracted and concentrated The effect of this extract on urine flow in a hydrated, ethanol-anesthetized rat is the usual method employed Effects of the extract are compared with effects of known amounts of vasopressin

Vasopressin release is controlled by the osmolality of the plasma and by the volume of the blood (Fig 7 25) Osmolality is sensed by osmoreceptors within the hypothalamus The localization of the osmoreceptors to the brain was first demonstrated by Verney (1947) Intracarotid injection of hypertonic saline was found to cause antidiuresis in dogs while similar volumes of hypertonic saline injected into the peripheral circulation had no effect An increase of 1 to 2 m Osm /kg is sufficient to raise vasopressin secretion provided that the osmolality is above 290 m Osm /kg Whether the osmoreceptors are the same cells that synthesize and secrete vasopressin is uncertain, they may be different cells When plasma osmolality increases, vasopressin is secreted

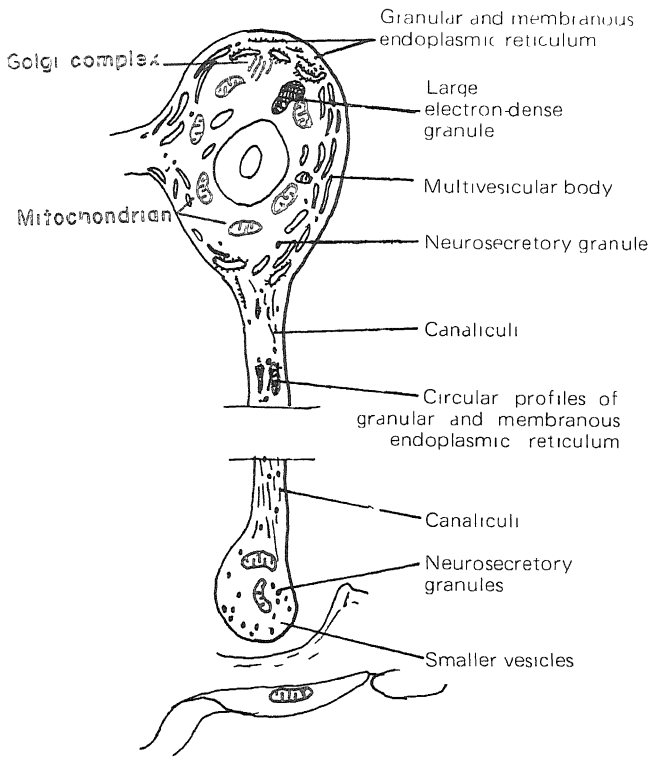


Fig 7 24 Diagrammatic representation of neuronal cell body in hypothalamus and nerve ending in infundibular process (From Sloper 1966)

Table 7 3
Neurohypophysial hormones of known chemical constitution and their distribution

Name	Chemical Constitution									Phyletic Distribution
	1	2	3	4	5	6	7	8	9	
Oxytocin	Cys	Tyr	Ile	Glu (NH_2)	Asp (NH_2)	Cys	Pro	Leu	Gly (NH_2)	Mammals, birds
Vasotocin	Cys	Tyr	Ile	Glu (NH_2)	Asp (NH_2)	Cys	Pro	Arg	Gly (NH_2)	All vertebrates except mammals
Arginine vasopressin	Cys	Tyr	Phe	Glu (NH_2)	Asp (NH_2)	Cys	Pro	Arg	Gly (NH_2)	Mammals except pig
Lysine vasopressin	Cys	Tyr	Phe	Glu (NH_2)	Asp (NH_2)	Cys	Pro	Lys	Gly (NH_2)	Pig

From Heller, 1966

and water reabsorption by the kidney is increased As a result, osmolality is restored to the initial value

Hemorrhage amounting to as little as 10% of blood volume results in an inhibition of water diuresis (Rydin and Verney, 1938) This response is due to an increase in the plasma vasopressin level (Fig 7 26) The response depends upon the degree of blood loss When blood loss approaches 50% of blood volume, vasopressin levels reach 600 μU /ml The release of vasopressin in response to hemorrhage is impaired in animals whose vagus nerves are cut Vagotomy interrupts the afferent pathway

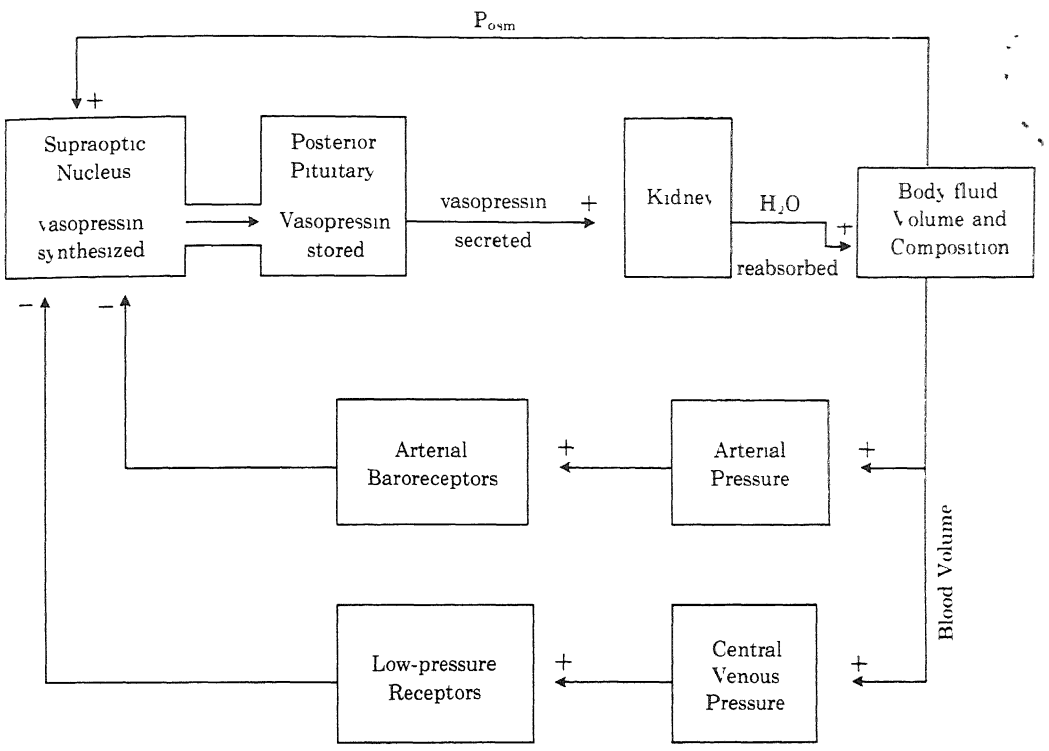


Fig 7 25 Control of vasopressin secretion by plasma osmolality and blood volume + increase or stimulation − decrease or inhibition (Modified from Share 1969)

from low-pressure receptors in the left atrium and perhaps other portions of the low-pressure side of the circulation Stretching of these receptors by distension of the left atrium generates afferent impulses which inhibit vasopressin release (Baisset and Montastruc, 1957, Share, 1965, Johnson et al, 1967) During hemorrhage, atrial pressure falls due to a drop in central venous pressure and fewer impulses arise from the low pressure receptors As a result, inhibition of vasopressin release is relieved and hormone appears in the blood

Carotid sinus baroreceptors are involved in control of vasopressin release When these receptors are stretched by normal or elevated blood pressure, impulses arise that are inhibitory to vasopressin release During hemorrhage, arterial pressure falls and fewer inhibitory impulses arise from the arterial receptors These receptors are probably responsible for the increase in plasma vasopressin seen in the vagotomized dog subjected to a large blood loss (Fig 7 26)

Interaction of the atrial and arterial pressure receptors in control of vasopressin release was suggested by Gauer and Henry (1963) In the normal animal, the first receptors to sense a change in blood volume are in the low-pressure system As hemorrhage becomes more marked, arterial blood pressure also falls and the arterial receptors participate in the increase in vasopressin secretion

Effects of vasopressin appear to be mediated by changes in intracellular levels of cyclic adenosine monophosphate (cyclic AMP) (Robinson et al, 1968) Addition of vasopressin increases cyclic-AMP levels while addition of cyclic-AMP mimics effects of vasopressin on water and sodium transport In addition, vasopressin stimulates the activity of adenyl cyclase found in preparation of kidney cell membranes The mechanism of action of vasopressin on the kidney is described elsewhere in this volume

CONTROL OF OXYTOCIN SECRETION

The major stimuli for release of oxytocin are *a*) suckling, and *b*) distension of the cervix and vagina (Cross, 1966) Increased milk ejection following oxytocin administration does not involve increased milk secretion but rather ejection of milk already present within the

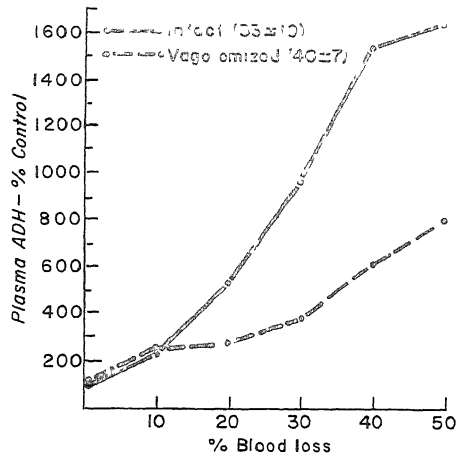


Fig 7 26 Effects of successive bleedings of 8 ml/kg body weight on the plasma vasopressin concentration in intact and vagotomized dogs The dogs were bled at 5-min intervals Blood samples for the determination of vasopressin were taken 3 min after the midpoint of each hemorrhage The plasma vasopressin concentration on the ordinate is expressed as per cent of the concentration of vasopressin in the plasma just prior to the first hemorrhage The abscissa is the degree of blood loss expressed as per cent of the initial blood volume which was taken to be 8% of the body weight The figures in parentheses are the concentration of vasopressin in microunits per milliliter of plasma in the initial blood samples (mean ± S E) (From Share 1969)

gland Ejection depends upon contraction of the myoepithelium present in the alveoli and lactiferous ducts Oxytocin levels have been measured in jugular vein blood before, during, and after suckling or milking and have been found to increase from about 5 μU /ml before suckling to 20 to 100 μU /ml during suckling Injection of 50 mU of oxytocin duplicated the normal milk-ejection response to suckling

The role of oxytocin in normal labor is still unclear Labor is a complicated process involving softening and dilation of the cervix, loosening of pelvic ligaments and contractions of abdominal muscles Oxytocin has no effect on these events On the other hand, blood levels of oxytocin in ewes were found to increase from about 3 μU /ml during pregnancy to 3,000 μU /ml during delivery of the lamb (Fitzpatrick, 1961 a,b) Similarly oxytocic activity of human systemic blood increased from 50 μU /ml during pregnancy to 150 μU /ml in the second stage of labor (Caldeyro-Barcia, 1961) Lesions that interrupt the supraopticohypophyseal tract in the anterior portion of the median eminence grossly impair labor (Fisher et al, 1938 b) Oxytocin, then, appears to serve an important but not vital role in labor It stimulates the frequency and duration of contractions in the uterine myometrium The hormone frequently is given by obstetricians to induce or intensify uterine contractions

Plasma levels of oxytocin are about the same in men as women, but the function of the hormone is uncertain Acceleration of sperm transport to the epididymis has been reported in rabbits

DISORDERS OF POSTERIOR PITUITARY FUNCTION

Diabetes insipidus is a condition in which large quantities of urine of very low specific gravity, 1 002 to 1 006, and low chloride content are excreted In an ordinary case the daily output of urine is 4 or 5 liters, but daily amounts several times these figures have been reported A corresponding increase in fluid intake occurs and thirst is often intense The condition frequently accompanies tumors of the pituitary or hypothalamic region, and is essentially due to lack of posterior pituitary hormone Ranson and his colleagues (Fisher et al , 1938a) were able to produce a similar picture in cats and monkeys by placing bilateral electrolytic lesions in the supraopticohypophysial tract in the hypothalamus Their cats developed a permanent polyuria of 300 to 500 ml of urine per day, which could be reduced to normal levels by replacement therapy with posterior pituitary extracts It may be pointed out that animals deprived of the pars distalis of the adenohypophysis, as well as the neurohypophysis, develop only a slight polyuria if any at all The pars distalis exerts a general diuretic effect which is necessary for the development of a marked polyuria This may be observed in human patients with an expanding lesion in the sella turcica Destruction of the neurohypophysis results in diabetes insipidus, but further development of the lesion may destroy the pars distalis with amelioration of the diabetes

Administration of vasopressin is the only available means for control of diabetes insipidus, but the hormone is not effective in some cases These individuals show a familial tendency to the disease which is thought to be due to an inherent defect of the renal tubules with respect to renal reabsorption of water

4 / CONTROL OF THYROID FUNCTION

The thyroid gland, as a distinct anatomical entity, has been known for at least 300 years, but its relationship to the clinical states we know as hyper- and hypothyroidism was not recognized until the latter part of the 19th century. At this time many of the features of cretinism and adult myxedema were noted to resemble changes which appeared after surgical removal of the thyroid. This suggested to physicians that the thyroid gland may secrete a humoral substance which is essential for the maintenance of well-being. A short time later Baumann discovered the presence of large amounts of iodine in the thyroid. Most of this was found to be incorporated within a protein which upon hydrolysis yielded a material that was effective in relieving the symptoms of myxedema. The physiologically active substance was isolated and chemically defined by Kendall (1915) who gave it the name thyroxin. In 1952, the second of the two presently known thyroid hormones, triiodothyronine, was discovered in thyroid glands by two English workers, Gross and Pitt-Rivers.

These two hormones, which are generally referred to as thyroxin or T_4 (3,5,3′,5′-tetraiodothyronine) and triiodothyronine or T_3 (3,5,3′-triiodothyronine), are both derived from the amino acid tyrosine, and are unique in that their biological activities are dependent on the presence of iodine, a trace element (Fig. 7-27). This particular structural feature has greatly aided our understanding of thyroid physiology and biochemistry. Since the radioactive isotope, ^{131}I, is incorporated readily into the hormones during their biosynthesis in the thyroid, it has been possible to study many aspects of the formation, distribution and metabolism of the thyroid hormones in a relatively straightforward way by measuring the isotope in vivo or in tissue samples.

HISTOLOGY

The thyroid tissue is composed of cuboidal epithelial cells arranged in a single layer around spaces roughly spherical in shape. These spaces variously known as follicles, vesicles, acini or alveoli contain a homogenous gelatinous material, the colloid substance which is the stored secretion of the gland. Connective tissue fibers support the alveolar walls and form septa which divide the gland into smaller masses. The cells lining the alveoli contain numerous mitochondria and a well-defined Golgi apparatus. Studies with the electron microscope have shown that the luminal or apical borders of these cells possess microvilli, which are characteristic features of cells engaged in absorption or secretion. When the gland becomes active, the Golgi apparatus hypertrophies and droplets of colloid appear in its proximity. Other histological changes typical of increased activity in the thyroid are increase in the epithelial cell height, increase in mitotic figures, decrease in amount of colloid in the follicles (with a tendency to become basophilic rather than acidophilic in staining properties), and an increase in vascularity. The maximal normal weight of the human thyroid is, according to Marine, from 20 to 35 g or around 0.4 g/kg of body weight. The thyroid tissue of the early fetus consists of masses of epithelial cells showing little or no arrangement into follicles. The latter appear about the 3rd month in the human but are small and contain little colloid. Functional activity first appears in the fetal gland at about this time. Concentration of inorganic iodide is demonstrable before the appearance of colloid, but organic binding (formation of iodinated amino acids) is not observed until colloid is demonstrable histologically.

Synthesis, Storage and Secretion of Thyroid Hormones

Iodine is a relatively scarce element in the earth's crust and animals have had to develop extremely efficient means to extract enough for biological purposes. Ultimately, all of the iodine taken into the body is derived from soil

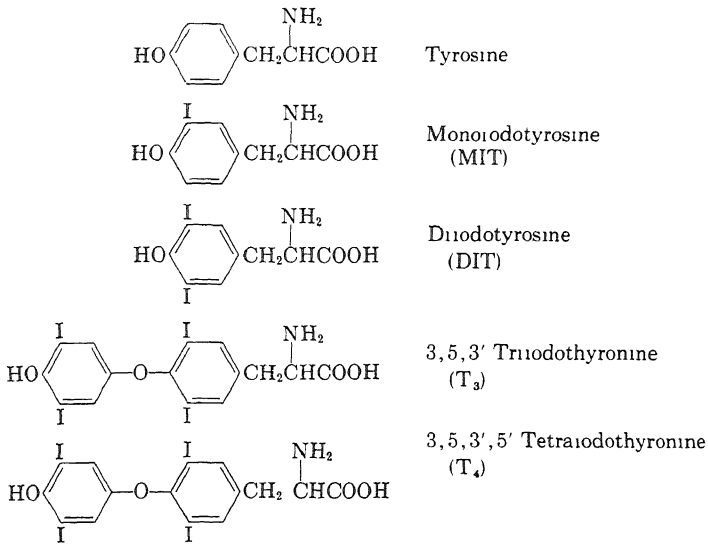


Fig 7 27 Structure of tyrosine and its important iodinated derivatives

and water and it is the content of iodine in these natural sources which determines the amount normally ingested Iodine taken in with food in organic form is released as iodide (I^-) and it is absorbed into plasma in this form (Fig 7 28) The major route of excretion of iodide is urine although small amounts are lost in the feces, mainly in organic form

IODIDE TRAPPING (Halmi, 1964, Wolff 1964)

Nearly one-half of the total body iodine is normally sequestered in the thyroid gland This is a truly remarkable accomplishment when one considers the extremely small amounts of iodine present in plasma and in the water and food that are consumed The ability of the thyroid cell to take up iodide from plasma and concentrate it is a prerequisite for the optimal synthesis of thyroid hormone In geographic areas where the content of iodine in soil and water is low, iodide trapping efficiency by the thyroid may become greatly increased In laboratory animals, such as the rat, the ratio of thyroid to serum iodide (T/S ratio) often reaches values of 100 when iodine depletion is rigorously maintained by the combination of a low iodine diet and the administration of propylthiouracil, a drug that blocks synthesis of thyroid hormone

The trapping of iodide against an electrochemical gradient, like active sodium transport, has been shown to require a steady input of energy from adenosine triphosphate (ATP) Transport is strongly inhibited under anaerobic conditions or following the administration of agents such as dinitrophenol which reduce ATP by uncoupling oxidative phosphorylation Inhibition of Na-K ATPase by the cardiac glycoside, ouabain, also interferes with iodide transport in thyroid slices, and the suggestion has been made that trapping may be linked in some way to cation transport

As in other carrier-mediated transport systems, the trapping of iodide can be inhibited competitively by other ions The strongest inhibitory effects are exerted by univalent anions of similar ionic size and shape Thiocyanate, perchlorate and nitrate ions are among the common ones which exhibit inhibitory properties Other halides such as fluoride and chloride do not react, although the heaviest halide, astatide, which is naturally radioactive, is accumulated by the thyroid to about the same extent as iodide The blocking effect of thiocyanate and perchlorate on iodide uptake by the thyroid will reduce the synthesis and subsequent release of thyroid hormones and, for this reason, both agents have been employed clinically in the treatment of hyperthyroidism

Although iodide trapping is a characteristic feature of the thyroid, it is not limited to this organ Other tissues such as the salivary glands, intestinal mucosa, mammary gland, the ciliary body and choroid plexus concentrate iodide to some extent and may even form small amounts of monoiodotyrosine (MIT), but

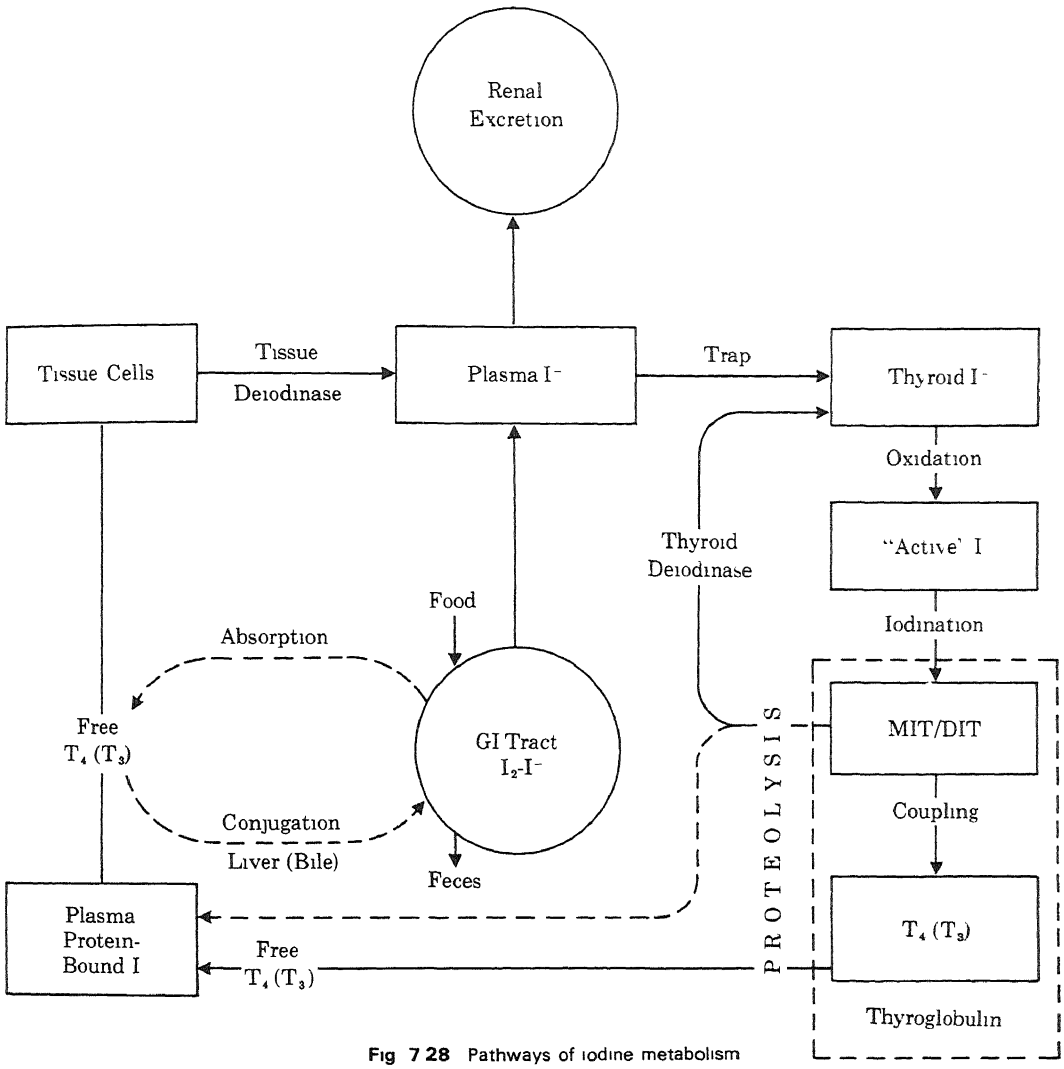


Fig 7 28 Pathways of iodine metabolism

these tissues do not possess the ability to synthesize thyroid hormones, and thus the total quantity of iodine stored is negligible It is of interest that iodide transport in these tissues can be inhibited competitively by the same anions that are effective in the thyroid However, trapping is not regulated according to the need for thyroid hormone as is the case in the thyroid gland

The mechanism of regulation of iodide transport in the thyroid is not clearly understood but one of the factors which may be involved is an adaptive response to thyroid-stimulating hormone (TSH) Iodide transport is greatly reduced in hypophysectomized animals and can be stimulated by the administration of TSH in vivo The response, however, is achieved only after several hours of treatment and attempts to stimulate iodide transport in vitro with TSH have been unsuccessful Thus, the effect appears to be indirect and a number of intermediate steps are undoubtedly involved in its regulation The notion has also been advanced that iodide transport is to some degree self-regulating As presently viewed, this regulation involves the inhibition of transport by some intracellular iodine product which as yet remains unidentified

ORGANIFICATION (Pitt-Rivers and Cavalieri, 1964, Rall et al 1964 Stanbury, 1966, Murray and McGuirr 1964 Greer et al, 1964, Taurog, 1970)

The second phase of thyroid hormone biosynthesis represents the iodination of tyrosine

residues in peptide-linkage within the intact thyroglobulin molecule a process known as organification or organic binding Since thyroglobulin is stored within the acinar space of the follicle iodide which is transported across the membrane at the base of the cell must move through the cell to the opposite or apical pole in order to reach the site of organification This apparently happens very quickly since in radioautographic studies in rats the first appearance of ^{131}I is in the colloid adjacent to the epithelial membrane at the apical pole Monoiodotyrosine (MIT) can be detected shortly after the injection of ^{131}I The formation of diiodotyrosine (DIT) and the coupling of the iodotyrosines to form T_4 and T_3 occur somewhat later

The mechanism of thyroglobulin iodination is not clearly understood but it is probable that most of the reactions involved are enzymatic in nature It is generally agreed that iodide is first oxidized at a site on the cell membrane adjacent to the acinus Biochemical and histochemical studies have revealed the presence of an enzyme, peroxidase, and it is also known that the organic binding of ^{131}I is strongly inhibited by the addition of catalase to preparations of isolated thyroid cells While these findings strongly suggest that hydrogen peroxide is an intermediate, the oxidation product of iodide has not been identified with certainty The iodination reaction, which involves the incorporation of oxidized or "active" iodine into the phenolic ring of tyrosine, is probably enzymatic as well and may represent an intrinsic function of the same peroxidase system

The final step in organification involves the coupling or condensing of two iodotyrosines, which are in peptide linkage within the thyroglobulin molecule to form the iodothyronines, T_3 and T_4 The reaction transfers the phenolic group from one iodotyrosine to another and splits out alanine which remains in peptide linkage The amount of T_4, as opposed to T_3, depends to some extent on the relative amounts of MIT and DIT present With extensive iodination of thyroglobulin, the ratio of DIT to MIT is increased and hence the yield of T_4 would be proportionately greater In man, as in a number of other animals, T_4 is the predominant iodothyronine It should be mentioned, however, that the much smaller quantity of T_3 does not necessarily lessen its physiological importance since the biological activity of T_3 is considerably more intense than that of T_4 Whether by design or by accident, this quantitative difference in effectiveness between T_3 and T_4 is useful from the standpoint of iodine conservation In deficiency states, when iodine must be efficiently reutilized, the proportion of T_3 increases, thus adding further protection against hormone deficiency

A number of cases have been reported in which spontaneous defects in coupling have resulted in goiter and hypothyroidism Only traces of T_4 and T_3 were found in these individuals although ^{131}I was rapidly taken up by the thyroids and incorporated into DIT and MIT which, in a few instances, could be detected chromatographically in the plasma

Blocking of Organification A number of organic compounds, including thiourea, propylthiouracil, the sulfonamide drugs, and mercaptoimidazoles, are known to interfere with organification and to cause goiter Owing to their effectiveness in blocking thyroid hormone synthesis, a number of these chemicals, especially propylthiouracil, and one of the mercaptoimidazoles, Tapazole, have been widely used clinically in the treatment of hyperthyroidism Naturally-occurring goitrogens, which exert similar effects on the thyroid, are found in the seeds of many plants of the cabbage family and in the rutabaga and turnip

The mode of action of these compounds has not been established but it is clear that all steps in organification are affected The coupling reaction is strongly inhibited, DIT and MIT formation are inhibited also although small amounts of MIT continue to be formed in the presence of a high concentration of blocking agent These agents exert no inhibitory effect on iodide trapping and in fact very high T/S ratios of iodide are usually observed following their administration This may result in part from the high levels of endogenous TSH which follow the reduction of thyroid hormone secretion

THYROID HORMONE STORAGE AND RELEASE (Rall et al, 1964, Wollman, 1969, Murray and McGuirr, 1964)

Thyroid hormone is stored within the gland as iodothyronines incorporated in peptide linkage within the protein, thyroglobulin The

latter is a protein of about 680,000 molecular weight which is synthesized by the epithelial cells and then secreted prior to its iodination into the lumen or acinus of the follicle where it accumulates as a material known grossly as colloid Since the acinar space is separated from the extracellular space by a layer of epithelial cells, the high molecular weight of the colloid serves to keep the osmotic pressure and, hence, the volume of the colloid space within reasonable limits

It is evident from the foregoing that the problem of hormone release becomes one of transporting the iodinated thyroglobulin back into the epithelial cell and digesting it The series of events which encompass this process have been dramatically unfolded with the use of the electron microscope Experimentally, as early as 5 min after TSH administration, the microvillous border of the epithelium adjacent to the colloid becomes extensively folded and cytoplasmic projections move out into the colloid Small vesicles, known as colloid droplets, appear within the cytoplasm This endocytic process resembles phagocytosis closely and is one of the earliest manifestations of TSH action As the vesicles move toward the basal membrane of the cell, they fuse with lysosomes which contain proteolytic enzymes The proteolysis of thyroglobulin and subsequent release of free T_4 and T_3 is presumed to take place within these vesicles Although this process can be visualized best under maximal TSH stimulation, it is thought to represent the normal route of thyroglobulin digestion and hormone release under normal physiological conditions

Thyroid deiodinase. In addition to its content of T_4 and T_3, thyroglobulin normally contains variable quantities of MIT and DIT which remain uncoupled These iodotyrosines are freed as a result of thyroglobulin digestion and, except for small amounts which escape into the circulation, they are rapidly deiodinated by a potent thyroid deiodinase This important function tends to conserve thyroidal iodide for reutilization in iodinating new thyroglobulin There are on record cases of hypothyroidism in which there is a genetic absence of thyroid deiodinase From this clinical evidence it may be presumed that without a mechanism for conserving thyroidal iodide the gland would lose more iodide than it could regain by augmented iodide trapping. The intracellular pool of iodide generated by this deiodination reaction does not mix readily with the transport pool of iodide and is thus portected against loss from the cell at sites of iodide exchange with the plasma This pool has been termed the second iodide pool as distinguished from the first or transport pool

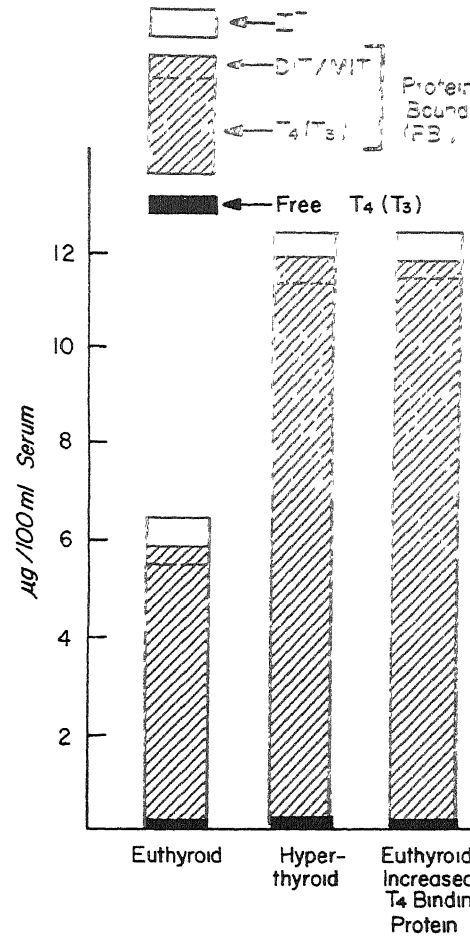


Fig 7 29 Iodine fractions in plasma

Distribution and Metabolism of Thyroid Hormones[1]

PLASMA IODINE

The iodine content of plasma exists in a number of discrete fractions and the analysis of these fractions in one way or another has been used extensively in diagnostic tests for hyper- and hypothyroidism (Fig 7 29) Of the

[1] Rall et al , 1964, Robbins and Rall, 1960, Tata, 1964, Sterling, 1964 and 1970

total amount of iodine in plasma, which in a healthy individual averages about 6 μg/100 ml, 90% represents iodine in organic form and the remainder is free iodide Of the organic fractions, more than 95% is bound to protein Since these protein-bound fractions, collectively called protein-bound iodine or PBI, largely comprise bound T_4 and T_3, the PBI has been used extensively as a diagnostic test for hyper- and hypothyroidism The normal range for the PBI is 4 to 8 μg/100 ml of serum, and values above and below this range are consistent with the diagnosis of hyper- and hypothyroidism, respectively Protein-bound T_4 and T_3 can be separated from iodide and small amounts of MIT and DIT by extracting the serum proteins with butanol and then washing the butanol with alkali This method has been used to improve the diagnostic capability of the PBI in instances, for example, where the PBI may be falsely elevated after the ingestion of large amounts of iodide

From the standpoint of the physiological effects of the thyroid hormones on body tissues, the most important fraction is the very small amount of free T_4 and T_3, estimated to be in the range of 5×10^{-11} M The quantity is too small to measure by direct chemical means, but it has been quantitatively assessed after equilibrium dialysis of plasma Most of it consists of T_4, but small amounts of T_3 are also present This free fraction is thought by most investigators to represent the physiologically-active fraction of circulating thyroid hormone and to correlate with the clinical states of hyper- and hypothyroidism

THYROXIN BINDING PROTEIN (TBP)

The small fraction of free thyroxin in serum is in equilibrium with a much larger protein-bound fraction Most of the bound thyroxin migrates during electrophoresis with a protein located between α_1 and α_2 globulin A smaller portion is bound to a prealbumin fraction In general terms, the quantity of T_4 that is bound can be predicted from mass action equations if the total number of sites that are available for binding are known In clinical situations, such as pregnancy or after estrogen administration, where the TBP is known to increase, the amount of thyroxin bound will also increase This might cause a significant reduction in the amount of free thyroxin were it not for the efficient operation of the thyroid-pituitary servomechanism which increases the output of thyroxin from the thyroid gland to maintain a physiological level of free thyroxin (see below) It should be pointed out that when the TBP is elevated, the PBI is also increased It is for this reason, among others, that the PBI has been largely supplanted by improved methods of diagnosis

FATE OF THYROID HORMONES

Free T_4 and T_3 in plasma are thought to be in equilibrium with free iodothyronines within tissue cells, and there is considerable indirect evidence suggesting that the cell membrane does not play an active role in the transport of the free hormone from the vascular to the tissue compartments, but rather the hormones gain access to the intracellular compartment by simple passive diffusion In ^{131}I label studies, the iodothyronines have been found to bind firmly but reversibly to subcellular constituents such as the nuclei, mitochondria, and the rough endoplasmic reticulum Owing to the mass of skeletal muscle, the largest fraction of tracer doses of T_4 and T_3 can be recovered from this tissue However, the most intense binding and highest rates of turnover are found in the liver It is of interest that virtually no intracellular binding of the thyroid hormones occurs in the brain or the gonads, tissues which are also unresponsive to the direct metabolic effects of these hormones

The iodothyronines are known to undergo at least three major types of reactions according to the tissue involved First, an appreciable fraction is taken up by the liver and conjugated The conjugated products appear in the bile largely as the β-glucuronide The product may undergo hydrolysis in the intestine followed by reabsorption of the free hormone This circuit from blood to liver to the intestine and back to blood again is known as the enterohepatic circulation A certain fraction of the β-glucuronide fails to be reabsorbed and is excreted in the feces

Like their amino acid cousins, the iodothyronines also undergo oxidative deamination and transamination These reactions occur both in the kidney and liver but are more prominent in the former and represent the major metabolic pathways for T_4 and T_3 in this tissue

The final and most important metabolic fate of the thyroid hormones is their deiodination

Eventually, iodine is irreversibly lost from all of the hormones secreted by the thyroid except for that fraction which is excreted intact in the feces The deiodination is accomplished by a tissue deiodinase activity which differs from thyroid deiodinase in that it has a high specificity for T_4 and T_3 Once iodide is liberated it may be retrapped by the thyroid or excreted in urine An interesting question may be raised as to whether T_3 is generated within tissue cells by the partial deiodination of T_4 While most investigators have felt that any T_3 present in tissue cells gets there by inward transport from the free T_3 pool in plasma, there is now evidence that an appreciable fraction of circulating T_3 is formed by the extrathyroidal deiodination of T_4

Control of Thyroid Function[2]

The thyroid gland is directly under the control of the anterior pituitary and the loss of pituitary function by hypophysectomy or disease, such as postpartum necrosis or tumor, results in striking involutional changes in the thyroid These changes include flattening of the follicular epithelial cells and decreases in rates of iodide uptake, hormone biosynthesis and release The PBI is often low and patients exhibit a low basal metabolic rate (BMR) and complain of extreme intolerance to cold All these features are directly attributable to the loss of thyroid-stimulating hormone (TSH) Under appropriate conditions, it is possible to reverse these changes and completely restore thyroid function It should be emphasized at this point that while thyroid function is greatly reduced in the absence of TSH, it does not entirely disappear and there remains a low intrinsic basal activity Thus, TSH does not appear to bring into existence any of the essential thyroid processes, but it simply accelerates their rates

THYROID-STIMULATING HORMONE TSH) (Bates and Condliffe, 1966 Odell et al, 1967, Field, 1968, Freinkel 1964)

TSH is a glycoprotein secreted by basophilic staining cells of the anterior pituitary It has a molecular weight of about 30,000 and has been purified sufficiently to permit the development of an immunoassay for endogenous TSH

Effects of TSH. The action of TSH is limited to the thyroid gland, but in this tissue the manifestations of its activity are numerous and they encompass nearly all phases of cellular function In general the effects are stimulatory and appear to be directed toward speeding-up the intrinsic processes that are evident in glands from hypophysectomized animals

For purposes of description, the effects of TSH may be divided into two major categories The first is concerned with iodine metabolism and includes all the steps involved in the transport of iodide from the plasma into the cell, its organic binding and the events leading to the hormone release Among the earliest effects of TSH is the stimulation of hormone release As mentioned earlier, the addition of TSH to thyroid slices or isolated cells is followed within minutes by the appearance of colloid droplets near the apical membrane of the cell This process which results in the secretion of T_4 and T_3 can be observed in the absence of any evident stimulation to organification or iodine trapping It is clear from other types of experiments however, that iodine binding is stimulated by TSH This effect is thought to be an independent response since it may be elicited under conditions where the trapping of iodide is blocked The stimulation of iodide trapping by TSH is the slowest of the responses and has not been directly observed in isolated thyroid tissue

Included in the second category are a large number of responses relating to energy metabolism, synthetic processes and cell growth Increases in oxygen utilization, glucose uptake, and oxidation are early responses to TSH Nicotinamide adenine dinucleotide phosphate (NADP) levels also rise in association with the stimulus to glucose oxidation While this alteration has given rise to considerable speculation, its significance has not been established Other early effects are the stimulation of ^{32}P incorporation into phospholipids and an increase in net Na^+ and water uptake Effects relating to cell growth are somewhat slower to appear These include the stimulation of protein and ribonucleic acid (RNA) synthesis, an increase in mitotic activity and an increase in the height of the epithelial cells Along with these changes, the gland often shows increased vascularity

Attempts to explain this diversity of effects by a single site of action has been a frustrating

[2] Purves, 1964, Dobyns, 1964, Brown-Grant, 1966

experience for many investigators in this field, especially since it has been shown that TSH does not enter the cell but probably is bound to receptor sites on the cell membrane As a result, TSH is spatially separated from the many locations of functions that it controls The answer to this problem may be the second messenger, cyclic AMP TSH has been shown to stimulate the adenyl cyclase activity in isolated thyroid cell membranes and to increase tissue levels of cyclic adenosine monophosphate (cyclic AMP) in intact cells Moreover, addition of the dibutyrl derivative of cyclic AMP, which penetrates the cell membrane, evokes many of the early responses to TSH, notably the stimulation of colloid droplet formation and glucose oxidation Whether all of the effects of TSH, particularly those relating to cell growth, are mediated to cyclic AMP yet remains to be established

PITUITARY-THYROID FUNCTIONAL RELATIONSHIP (Fig 7 30) (Purves, 1964 Dobyns 1964, Brown-Grant 1966 Odell et al , 1967)

It has long been recognized that the release of TSH by the pituitary is strongly inhibited by T_4 and T_3 It is generally thought that the inhibitory response is sufficiently sensitive to physiological concentrations of hormone that any significant tendency for the level of free T_4 to deviate from a set point would be buffered by appropriate changes in the secretion of TSH This in a few words describes the basic mode of regulation of thyroid secretion by the pituitary In operation, it is identical to the type of negative feedback control or servomechanism found in many different control systems and serves to maintain the level of hormone in plasma at a constant value When the constant inhibitory play of free T_4 and T_3 on the pituitary is decreased or abolished following thyroidectomy or the administration of agents such as propylthiouracil which interfere with hormone secretion, TSH release increases The subsequent rise in plasma TSH can be measured in samples of plasma by immunoassay methods and will differentiate hypothyroidism resulting from a primary interference with thyroid hormone production in the thyroid from that arising secondarily as a consequence of the loss of TSH secretion by the pituitary There is some evidence, largely from studies of cold adaptation, that T_4 also exerts an inhibitory effect on TSH secretion via the hypothalamus This effect, however, does not play a prominent role in the feedback control of the thyroid

INFLUENCE OF THE HYPOTHALAMUS ON THE PITUITARY-THYROID RELATIONSHIP (Brown-Grant, 1966 Burgus and Guillemin 1970b)

The basic feedback relationship between the pituitary and thyroid can operate independently of the central nervous system, but it is clear from a number of physiological studies that higher centers do influence this relationship to a significant degree If the neurovascular link between the hypothalamus and pituitary is interrupted experimentally, either by transplanting the pituitary to another site such as the renal capsule or anterior chamber of the eye or by transecting the pituitary stalk and inserting an impermeable barrier to prevent growth of the portal vessels, the activity of the thyroid declines somewhat, but is still significantly higher than that found in hypophysectomized animals Although the amount of TSH that is secreted is smaller than that secreted when the pituitary is in its normal position, the secretion can be inhibited by T_4 Thus, the basic servomechanism remains intact, but it appears as though the set point has been adjusted downward One of the more striking features of animals with transplanted pituitaries is their failure to release TSH with cold stress as occurs in normal animals

An exciting recent development has been the chemical identification and synthesis of a small peptide serving as the humoral link between the hypothalamus and the anterior lobe of the pituitary The substance which had been recovered from hypothalamic extracts is called thyrotropin releasing factor or TRF It is thought to be synthesized by neurons in the hypothalamus and secreted at nerve endings in the median eminence where it is picked up by vessels of the hypophysial portal circulation and then carried to the anterior lobe There it acts to stimulate the release of TSH Although much has yet to be learned about the relationship of the hypothalamus to the anterior pituitary, it may be assumed that much of the regulatory activity that has just been discussed is a result of the interplay between the stimulatory

effect of TRF and inhibitory influence of T_4 and T_3 on TSH secretion by the anterior lobe It follows from this that the level of TRF stimulation determines the set point or "thermostatic setting" about which the pituitary-thyroid servomechanism operates There is already considerable evidence that the setting is increased in homeothermic animals when they are exposed to cold

THYROID REGULATION IN SOME COMMON SITUATIONS

Goiter. Any agent or condition which interferes directly with thyroid hormone production will promote increased TSH secretion by decreasing the inhibition of TSH release If the interference is prolonged, thyroid enlargement or goiter may result This is the usual mechanism of goiter formation associated with the administration of thyroid blocking agents such as propylthiouracil or thiocyanate, or following ingestion of naturally occurring goitrogens In all of these instances, the thyroid enlargement is dependent on continued TSH secretion since, in animal experiments, removal of the pituitary or suppression of TSH release by the administration of T_4 will prevent goiter formation

Thyroxin-Binding Proteins (TBP) It was mentioned earlier in this chapter that the physiological activity of thyroid hormone on tissues is more likely related to the free fraction in plasma than to the much larger amount bound to protein The quantity of binding-protein is known to increase considerably in such situations as pregnancy or following the administration of estrogens and may decrease where there are abnormal urinary losses of protein as in nephrosis It is of interest then to consider the consequences of changes in the level of binding proteins on the regulatory mechanisms As the level of binding protein rises for any reason, more T_4 will be bound and the free fraction of T_4 will tend to fall Owing to the operation of the pituitary-thyroid servomechanism, any reduction in free T_4 will evoke an increase in the rate of secretion by the thyroid This response will take place concurrently with the rise in TBP and when the rise ceases, many of the additional binding sites on the protein will be occupied by T_4, thus contributing to an elevation in the PBI

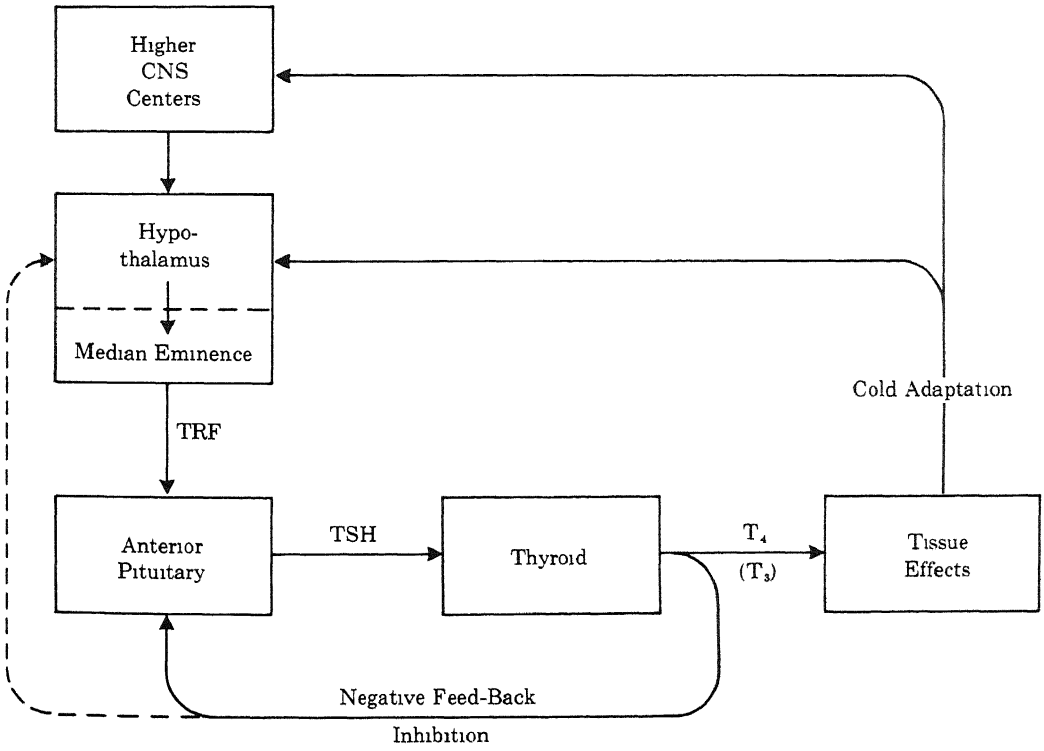


Fig 7 30 Scheme showing major regulatory influences on the thyroid

The concentration of free T_4, however, will remain the same since it is adjusted by the setting of the servomechanism Thus, the end result of the physiological adjustment to a rise in TBP is a larger pool of T_4, most of which is bound Hyperthyroidism does not occur in this situation since tissue effects of T_4 are related to the free fraction, not the amount bound

The opposite changes occur when the level of TBP falls, as for example following pregnancy An effect similar to an actual fall in the level of binding protein may be obtained by the administration of drugs such as salicylates which compete with thyroxin for the binding sites

Tissue Effects of Thyroid Hormone

The mechanism of action of the thyroid hormones at the tissue level is not known It is clear, however, that the iodothyronines T_4 and T_3 exert numerous effects on nearly every tissue of the body While the diversity of these effects seems endless and limited only by the imagination and energy of the investigator, most of the effects appear to participate in the regulation of two general types of processes *a*) growth and differentiation, and *b*) oxidative or energy metabolism Aside from purposes of discussion, there are reasons for believing that these two categories of responses may represent different actions of thyroid hormone It is known, for example, that T_4 will stimulate oxygen utilization in all homeothermic vertebrates, but little or no response is seen in poikilothermic animals On the other hand, metamorphosis in amphibia and other poikilothermic vertebrates is strongly dependent on the presence of a functioning thyroid The basis for such a differentiation may be more apparent than real, however, since investigators have now begun to recognize that the effect of thyroid hormone on oxygen utilization may depend on an alteration at the level of gene expression

GROWTH AND DIFFERENTIATION (Shallabarger, 1964)

Early studies have shown a clear dependence of metamorphosis in frogs on thyroid hormone The administration of thyroid extract to frog tadpoles results in rapid metamorphosis and the production of small but otherwise normally developed frogs Contrarily, thyroidectomized tadpoles fail to develop legs although general somatic growth continues Thus, very large larval forms are produced in this way Mammalian development is also influenced by thyroid hormone Hypophysectomy in very young rats treated with growth hormone will show general growth of bone, but striking abnormalities in the epiphyseal structure will appear While T_4 alone in these animals has little apparent effect on general somatic growth, it is required for the normal pattern of bone maturation Development of the central nervous system is also affected by the thyroid, a response which is thought to account for the mental retardation in cretinism where thyroid function is lacking early in life

CELLULAR METABOLIC PROCESSES (Barker 1964 Wolff and Wolff 1964

In 1895 Magnus-Levy observed a low basal oxygen consumption in hypothyroid patients and was able to reverse this defect by the administration of thyroid extract It has been shown subsequently that basal oxygen consumption in the range of 50% below normal to twice the normal rate is directly influenced by the level of circulating thyroid hormone Both T_3 and T_4 have qualitatively the same effect on tissue respiration and other metabolic processes, although there are quantitative differences that were alluded to earlier Investigation has shown that mole for mole the effect of T_3 on oxygen consumption is quicker to appear and gives a greater peak response than does T_4 However, since the total duration of effectiveness of T_3 is shorter than T_4, both drugs are thought to give about the same overall response

Alterations in oxidative metabolism are also evident in isolated tissues taken from hyper- and hypothyroid animals Large changes in O_2 utilization are seen in liver, muscle, kidney, heart, salivary glands, and the pancreas, while other tissues such as brain, spleen, lymph nodes, and the gonads, are apparently unresponsive to the hormone It is of interest to note, as was mentioned earlier, that labeled T_4 does not bind appreciably to subcellular particles from the latter group of tissues

The mode of action of thyroid hormone on oxygen consumption has been the subject of extensive investigation for many years In general, all oxidative processes are speeded up in

tissues from hyperthyroid animals The utilization of energy yielding substrates, such as glucose and fatty acids, are greatly increased but it is clear that the effect is not simply the result of an increase in the availability of these substrates for oxidation Some basic change has occurred in oxidative machinery of the cell In view of the similarity between hyperthyroidism and the hypermetabolism which results from the administration of uncouplers of oxidative phosphorylation, such as dinitrophenol, the notion has been put forth that thyroid hormone acts as an uncoupling agent It is true that some deficiency in ATP formation does occur, particularly in very severe hyperthyroidism, but in slight or moderate hyperthyroidism ATP levels may be normal or even increased Furthermore, since it is not possible to couple oxidative phosphorylation more tightly than one sees in euthyroid states, it is difficult to explain the decreased oxygen consumption in hypothyroidism Most of the evidence clearly points to the regulation of oxidative metabolism within the mitochondria, with or without effects on the coupling of oxidation to ATP formation

While the biochemical mechanism of the thyroid hormone-induced hypermetabolism is still not clear, it does seem probable that the effect is dependent on some fundamental alteration of the cell which involves new protein synthesis This alteration could represent enzyme protein, structural components or a cluster of changes It has long been known that the effects of T_4 and even the faster-acting T_3 on O_2 consumption are slow to appear, often requiring days of treatment before optimal increases have been established One may presume, of course, that other changes occurred much earlier since it is known that the hormones attach themselves to cellular sites soon after they penetrate the cell membrane Earlier effects have in fact been measured and one of the first recorded is an increase in the synthesis of nuclear RNA in livers of hypothyroid rats This is followed by the stimulation of protein synthesis in certain fractions of cell protein, including mitochondrial protein Of interest is the further finding that when protein synthesis is inhibited, the enhancement of O_2 consumption following T_4 fails to appear Such an apparent dependence of tissue respiration on protein synthesis in this instance does not imply that all metabolic effects of T_4 and T_3 are related to protein synthesis, but it does implicate one or more of the key steps leading to hypermetabolism

A number of other metabolic changes are seen in hyper- and hypothyroidism Owing to the need for large quantities of oxidizable substrate in hyperthyroidism, glycogen stores in liver may become rapidly depleted in the postabsorptive state and the breakdown of body protein, especially skeletal muscle, may be increased to provide carbon for gluconeogenesis The oxidation of fatty acids and cholesterol is increased in hyperthyroidism The opposite effect is seen in hypothyrodism where hypercholesterolemia is a prominent feature Because of this blood finding and the high incidence of angina and ischemic heart disease in hypothyroidism, considerable effort has been expended in synthesizing thyroid hormone analogues which will lower serum cholesterol without producing significant calorigenic effects Despite the prominent catabolic effect of thyroid hormone on protein metabolism that is evident clinically in severe hyperthyrodisim, thyroid hormone does exert significant anabolic effects on tissue and body growth The effect is clearly seen from the retardation of growth that occurs when adequate amounts of thyroid hormone are lacking during development

Physiological Aspects of Thyroid Disorders[3]

THYROID ^{131}I UPTAKE (Goolden 1964)

A variety of tests have been devised to measure thyroid function by the amount of ^{131}I-iodide taken up in the thyroid under specified conditions By and large these have proven to be useful not only in the diagnosis of thyroid disorders but in extending our knowledge of the basic functioning of the thyroid As is so often the case, the two objectives are inseparable since an intelligent understanding of disease requires a full knowledge of the working of the system in health Most of the tests have employed external counting of radioactivity over the thyroid region in the neck from 10 min to 48 hr after the test dose Since plasma iodide is cleared almost exclusively by the thyroid and the kidneys, the quantity of radioactivity excreted in the urine

[3] MacGregor, 1964, Morgans, 1964

will provide an indirect assessment of uptake Other modifications have included the measurement of ^{131}I in PBI after 24 to 48 hr as an assessment of hormone secretion The average range of values for the 24-hr thyroid uptake in normal subjects is 10 to 40% of the test dose Values above 50% are indicative of hyperthyroidism (Fig 7 31)

The ^{131}I uptake is influenced by many factors, some of which should be considered in the interpretation of clinical data The ingestion of iodides will reduce uptake by dilution of the label in plasma and by a direct, but not well understood, suppression of thyroid activity Thyroid blocking agents, of course, exert inhibitory effects Since the pool of free iodide in the thyroid is small compared with the pool of organified iodine, ^{131}I uptake in the presence of a block in organification will not be large The administration of perchlorate will discharge this pool of ^{131}I, however, but will not reduce any radioactivity that has been organically bound This method has proved useful in differentiating defects in organification from other forms of hypothyroidism

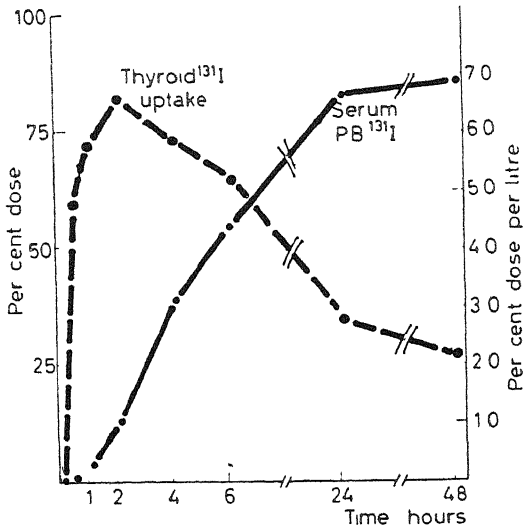


Fig 7 31 The accelerated turnover of iodine in thyrotoxicosis demonstrated by ^{131}I studies in a thyrotoxic woman following an intravenous dose of 25 μc The thyroidal uptake of ^{131}I was very rapid the thyroidal clearance rate being 302 ml/min/1 73 m ^{2} a spontaneous release of radioactivity beginning after 2 hr was reflected by a steady increase of protein-bound radioactivity in the serum at all times more than 90% of this PB ^{131}I was extractable with butanol chromatographic analyses confirmed the presence of iodothyronines from 2 hr onwards (From Murray and McGirr 1964)

The change in uptake after the test administration of TSH or T_4 (T_3) has on occasion proved to be extremely useful diagnostically In hypothyroidism of pituitary origin, the administration of TSH will greatly stimulate uptake Thyroxin or T_3 will suppress uptake by its inhibitory effect on TSH release In the hyperthyroidism of Graves' disease however uptake is not suppressable with T_3 and this test has been widely employed to differentiate borderline cases of hyperthyroidism of this kind from otherwise normal individuals with overly active thyroids

HYPERTHYROIDISM

This clinical condition results from excessive levels of thyroid hormone in plasma The term is a physiological designation and, hence, does not denote the source of the hormone Generally, the cause is hypersecretion by the thyroid, but an excessive oral intake of thyroxin will give the same result Hypersecretion by the thyroid commonly arises either from a diffuse hyperplasia of the thyroid or from hyperplastic nodules Rarely, it is associated with thyroid malignancy The first of these afflictions classically occurs in young women and is known in this country by the eponym, Graves' disease The cause of the diffuse hyperplasia has not been clearly established, but there is abundant evidence that it is not the result of an overproduction of TSH since plasma levels are usually within normal limits or even reduced somewhat More recent findings indicate the presence of a long-acting thyroid stimulator (LATS) in the plasma of many patients with this disorder LATS is a protein recovered from γ globulins which is thought to arise as part of an autoimmune response to tissue antigens It exhibits many of the properties of TSH but is clearly not of pituitary origin and would appear to be able to exert its effects independently of the normal thyroid control mechanisms Nodular hyperplasia is more common in older individuals and when associated with hyperthyroidism only portions of the gland may be overly active

Regardless of the cause of the excessive hormone secretion, the effects are about the same and consist of intolerance to heat, increased basal oxygen consumption (BMR), cutaneous vasodilatation and sweating, increased appetite despite weight loss, an in-

crease in the basal heart rate, a fine tremor of the hands, emotional irritability, and characteristic eye signs consisting of lid retraction, exophthalmos, and swelling of the circumorbital tissues The eye signs are more commonly seen in hyperthyroidism of diffuse hyperplasia than in cases of so-called "toxic nodule" Many of the signs and symptoms of hyperthyroidism closely simulate effects of epinephrine While there is no evidence that excessive secretion of epinephrine is a primary factor in hyperthyroidism, it is clear that the effects of these agents are strongly potentiated in this disorder Thus, efforts to reduce epinephrine secretion, by the use of ganglionic blocking agents, have been successfully applied in the treatment of severe thyrotoxicosis

The clinical diagnosis is usually confirmed by objective measurements in the laboratory Estimates of the BMR were widely used in the past but have been superseded by more reliable and convenient diagnostic procedures These include estimations of the amount of hormone in plasma by the PBI, the butanol-extractable iodine or the T_3 uptake test The last involves adding radioactive T_3 to the patient's plasma in vitro and measuring the uptake of T_3 by resin after incubation Assuming the level of binding protein to be normal, a high T_4 level results in a low concentration of unoccupied binding sites, little binding of T_3 and a high uptake by the resin As mentioned above, the uptake of ^{131}I by the thyroid is usually increased, but of more importance diagnostically, is its failure to be suppressed by a trial of T_3 The gland behaves as though it were autonomous and free from pituitary control

The treatment of hyperthyroidism consists of reducing the activity of the thyroid gland by removing it surgically, destroying thyroid tissue by radiation with ^{131}I or by the use of antithyroid drugs such as propylthiouracil, methyl mercaptoimidazole, or perchlorate The choice of treatment in any particular case must be carefully weighed after considering the many factors which bear on treatment The reader is referred to standard clinical textbooks for further discussion

HYPOTHYROIDISM

The consequences of thyroid hormone deficiency are greatly influenced by the age of the individual at the onset of the deficiency When it is present in infancy, treatment must be started promptly or dwarfism, retention of infantile body proportions, and mental retardation will invariably result A term that has been applied to this condition is cretinism Congenital hypothyroidism is most commonly caused by environmental iodine deficiency, although many cases of inherited enzyme defects in hormone production associated with goiter have been reported Adult hypothyroidism may be of pituitary origin, but it is usually the consequence of a primary thyroid disorder Loss or greatly reduced thyroid function may result from the administration of thyroid blocking agents, thyroidectomy or destruction of the gland by ^{131}I, or as the terminal stage in a disease process known as Hashimoto's disease or autoimmune thyroiditis The clinical picture of classical hypothyroidism is one of a gradual onset of lethargy, intolerance to cold, voice changes, slowing of mental faculties, and mental confusion There is a characteristic thickening of the skin of the lips, fingers, below the eyes, and of the legs The thickening is the result of the deposition of mucoprotein in the subcutaneous tissues, and when excessive, is called myxedema

The clinical diagnosis of hypothyroidism may be confirmed in the laboratory by the use of the same tests described under hyperthyroidism When hypothyroidism of pituitary origin is suspected, an increase in ^{131}I uptake by the thyroid after a trial of TSH administration may be helpful Hypothyroidism is treated by the oral administration of T_4 or crude thyroid extract The dose is adjusted after careful consideration of the patient's cardiac status since large initial doses may precipitate acute myocardial ischemia in patients with long standing hypothyroidism

5 / CONTROL OF CALCIUM HOMEOSTASIS BY PARATHYROID HORMONE, CALCITONIN, AND VITAMIN D

The parathyroid glands and thyroid C cells provide hormonal control for fine adjustment in extracellular calcium concentrations Many processes require ionic calcium, within narrow limits of concentration, to function properly Functions of calcium fall into two general categories *1*) maintenance of specific membrane permeabilities, and *2*) activation of enzymatic reactions Examples of calcium-requiring processes include acetylcholine release, neuromuscular excitability, membrane adenosine triphosphatase (ATPase) activities and actin-myosin interaction in muscle contraction Reviews of calcium homeostasis can be found in Copp (1969 and 1970)

The major factors involved in regulation of calcium homeostasis are diagrammed in Figure 7 32 Nonhormonal regulation of blood calcium involves changes in calcium influx through the intestine and excretion through the kidney that are brought about in following ways *1*) If calcium content of the diet is low, little is available for intestinal absorption, but the low amount increases the capacity of the intestine to take up calcium The effective amount of calcium in the diet can be lowered by certain foods such as some cereals and spinach These foods contain acids (phytic and oxalic) forming insoluble complexes with calcium and preventing its absorption *2*) Calcium absorption is changed by the amount of inorganic phosphate in the diet A low level of phosphate increases calcium uptake and a high level decreases it Work with experimental rat diets indicates an optimal ratio of 0 05% Ca/0 03% P Calcium and phosphate absorption are also influenced by parathyroid hormone as will be discussed later *3*) Absorption of calcium by the duodenum and upper jejunum is partly active and partly passive Active calcium transport has been known to require vitamin D But as seen in Figure 7 33, a vascularly perfused isolated intestinal system from a vitamin D deficient animal was not responsive to the vitamin itself when added to the perfusate However, in vivo injection of D_3 12 hr prior to the experiment increased calcium transport The preparation was much more responsive to 25-hydroxycholecalciferol (25-(OH)D_3), a metabolite of D_3 It has since been found that (1,25-(OH)$_2D_3$) is the active metabolite of D_3 (Omdahl et al, 1971) This compound is synthesized in the kidney (Fraser and Kodicek, 1970) and acts to induce synthesis of a specific calcium-binding protein in the intestinal mucosa cells (Wasserman and Taylor, 1966) and a Ca^{++}-dependent ATPase in the brush border Deoxyribonucleic acid (DNA) transcription is not required for the increased calcium transport, but is required for conversion of D_3 or 25-(OH)D_3 to 1,25-(OH)$_2D_3$, (Tanaka et al, 1971) It has been suggested that 1,25 (OH)$_2D_3$ be classified as a hormone controlling calcium metabolism *4*) Hypomagnesemia nearly always results in hypercalcemia, although the mechanism involved is not clear Magnesium itself may be regulated by parathyroid hormone In some instances, these divalent cations may substitute for each other, while in others, their effects are opposing In any case, magnesium modifies the effectiveness of calcium ions *5*) In addition to these effects on intestinal absorption, excretion of calcium through the kidneys can play an important role in nonhormonal control of hypercalcemia When plasma calcium increases above normal, urinary excretion increases rapidly

HORMONAL REGULATION

Two hormones are responsible for fine regulation of blood calcium level, parathyroid hormone (parathormone or PTH) from the parathyroid glands and calcitonin from the ultimobranchial tissue whether present as "C"

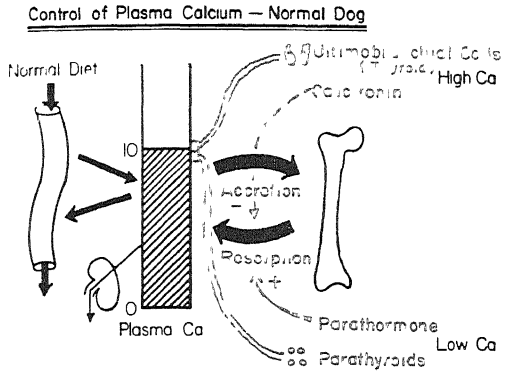


Fig 7 32 Diagram of factors involved in calcium homeostasis Intestinal absorption renal excretion and accretion and resorption of Ca^{++} by bone are the major factors involved Calcitonin produced by the C cells of ultimobranchial origin and parathyroid hormone produced by the chief cells of the parathyroid are major regulatory factors (From Copp 1968)

cells in mammalian thyroid or as a separate gland in other classes of vertebrates The two hormones work in a reciprocal manner to maintain blood calcium concentration near 10 mg/100 ml (5 mEq/liter), with only a small diurnal variation, in spite of large changes in intake Although an adult has a total of 1000 to 1200 g of calcium, more than 99% is unavailable in ionic form since it is crystallized in bone or complexed with protein or other anions in blood (Table 7 4) Quite obviously, blood calcium homeostasis must be regulated by sensitive mechanisms that can be rapidly initiated and terminated

PARATHYROID GLANDS

Parathyroid glands are found in all animals above the fishes In humans, there are usually two pairs, located close to the posterior surface of the thyroid in the neck The number and location may vary significantly, however Total weight for the four glands is about 120 mg Even though the parathyroids were known since 1909 to be associated with blood calcium and with tetany, it was not until 1959 that a fairly pure preparation of PTH was made (Aurbach, 1959) PTH is secreted by the chief cells, especially the dark-staining type of the parathyroid gland PTH is a single chain polypeptide with 83 amino acids and a molecular weight of approximately 9500 (Potts et al, 1968) The amino acid sequence of bovine hormone has been determined (Brewer and Ronan, 1970), and found to have 84 amino acids Only small amounts of human hormone are available for structural studies A portion of the amino terminal end of the molecule (35 amino acids) has significant biological activity and essentially full immunological activity

The overall physiological function of PTH is to maintain blood calcium at 10 mg/100 ml by providing more calcium if the blood level falls For example if an animal is parathyroidectomized, blood calcium will fall from 10 to 6 or 7 mg/100 ml Injection of PTH will re-

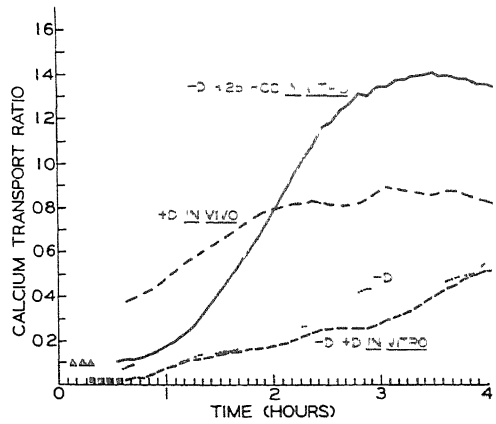


Fig 7 33 Response of vascularly perfused intestine to vitamin D and HCC (From DeLuca 1970)

Table 7 4
Forms of calcium and phosphate in normal plasma

	Conc (*mmoles/litre*)	Per Cent of Total
Calcium		
Free ions	1 16	47 5
Protein bound	1 14	46 0
$CaHPO_4$	0 04	1 6
$CaCit^-$	0 04	1 7
Unidentified complexes	0 07	3 2
Total	2 45	100 0
Phosphate		
Free HPO_4^-	0 50	43 0
Free $H_2PO_4^-$	0 11	10 0
Protein bound	0 14	12 0
$NaHPO_4^-$	0 33	29 0
$CaHPO_4^-$	0 04	3 0
$MgHPO_4^-$	0 03	3 0
Total	1 15	100 0

From Walser, 1961

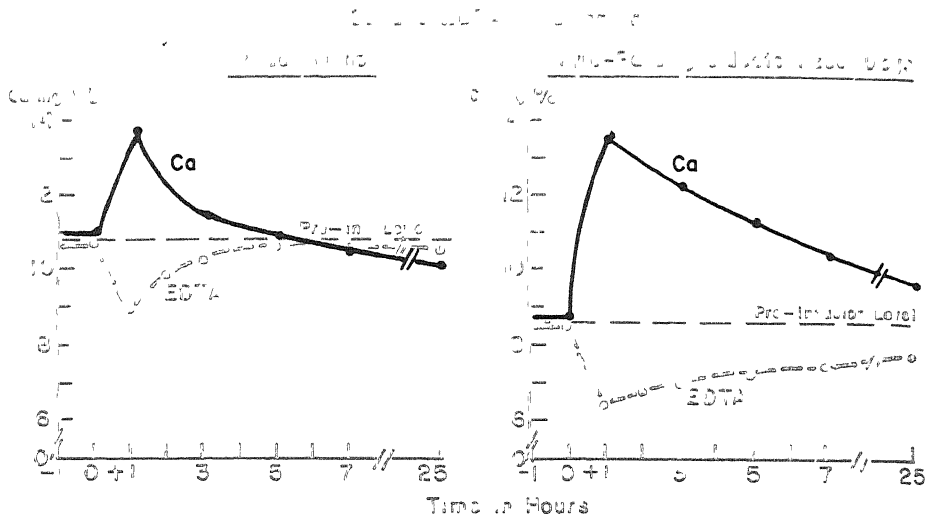


Fig 7 34 Serum calcium levels following calcium and EDTA infusions in normal and parathyroidectomized dogs (From Sanderson et al 1960 as adapted by Copp 1968)

turn the level to normal The parathyroidectomized or thyroparathyroidectomized animal (thyroid and parathyroid glands removed) cannot rapidly compensate for hypercalcemia or hypocalcemia Infusion of calcium or ethylenediaminetetraacetic acid (EDTA) (a calcium-binding agent), systemically, results in a more prolonged period of hyper- or hypocalcemia (Fig 7 34) The only known stimulus for secretion of PTH is low blood calcium This was demonstrated experimentally by perfusing the thyroid-parathyroid gland complex through the arterial supply with blood containing various levels of calcium or EDTA (Copp and Davidson, 1961) Although venous drainage from the glands remained intact, amounts of calcium or EDTA used in the perfusions were not enough to change systemic calcium levels Thus, any change found peripherally was due to hormonal action Using this technique, systemic calcium levels were found to rise if the glands were perfused by EDTA-containing blood, and to fall if high-calcium blood was used With the aid of a sensitive radioimmunoassay, it is possible to measure blood levels of PTH PTH secretion rate is found to be a linear function of plasma calcium concentration in the range of 6 to 12 mg./100 ml (Fig. 7 35) The half-life in blood is 20 min (Melick et al, 1965)

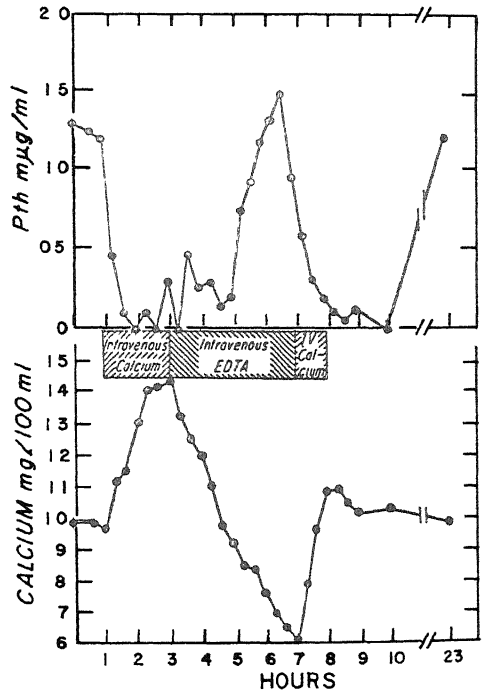


Fig 7 35 The effect of intravenous infusion of calcium and EDTA upon plasma calcium (*lower*) and plasma parathyroid hormone concentrations (*upper*) (From Care et al 1966)

ACTIONS OF PTH

Three sites of action are involved in the ability of PTH to raise plasma calcium concentration The first and probably most important site is bone, where PTH stimulates resorption and release of calcium The mineral in bone is mostly in the form of a complex crystal called hydroxyapatite, $[(Ca_3PO_4)_2]_3$,

$Ca(OH)_2$ Minerals are either part of this crystal or adsorbed to the surface of it A large surface area provided by the crystal arrangement facilitates mineral exchange between bone and blood Some calcium in bone will exchange with plasma calcium rather easily (exchangeable calcium) and some will do so slowly or at an immeasurable rate (stable or nonexchangeable calcium) PTH mobilizes bone calcium from two areas, osteolysis increases around large osteocytes and osteoclast activity increases (Figure 7 36) The result is resorption of bone, and in large part, it is the stable bone that is broken down In this process, the organic matrix of bone is dissolved as well, and hydroxyproline, a product of collagen dissolution, is excreted in the urine This substance forms the basis for assessment of the relative rate of bone dissolution and is used clinically for this purpose Release of calcium appears to occur in two stages *a*) a small amount is liberated in the first several hours, and *b*) a larger amount is liberated more slowly over a longer subsequent period (Rasmussen et al, 1964) Release of the larger amount requires RNA synthesis, since actinomycin D, an inhibitor of this process, will inhibit calcium release PTH has been shown to directly stimulate RNA synthesis in vivo and in vitro in bone (Steinberg and Nichols, 1968)

PTH appears to stimulate adenyl cyclase and to result in the formation of increased amounts of cyclic adenosine monophosphate (cyclic AMP) in bone (Fig 7 37) Calcitonin has only a small effect on cyclic levels and does not antagonize or enhance the effects of PTH

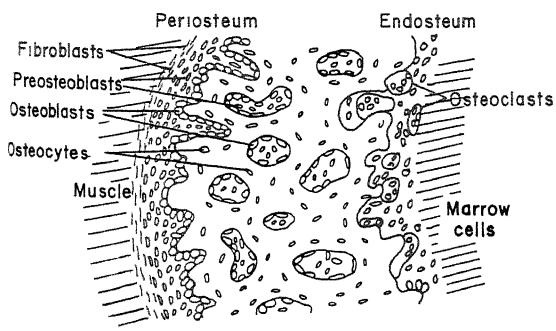


Fig 7 36 Diagrammatic representation of part of the cross section of bone wall from the mid shaft of the femur of a 10-day-old rabbit Layers of cells on the periosteal surface are associated with bone growth while cells on the endosteal surface are associated with bone resorption (From Owen and Bingham 1968)

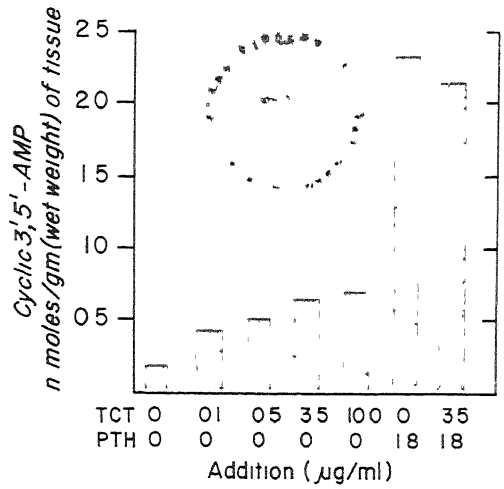


Fig 7 37 Effect of calcitonin (TCT) and parathyroid hormone on 3' 5' AMP in skeletal tissue during 5 min of exposure (From Chase and Aurbach 1970)

Addition of dibutyryl 3'-5'-AMP causes effects in vivo and in vitro similar to those recognized as physiological respones to PTH Parathyroid hormone and dibutyryl 3'-5'-AMP cause increased rates of glycolysis and increased synthesis and release of acid hydrolases, enzymes that are associated with lysosomes (Vaes, 1968a) Increased glycolysis results in production of acid that solubilizes bone mineral Acid hydrolases digest the organic matrix These factors account for release of calcium, phosphate and amino acids into the blood (Fig 7 38) The full effect of PTH on bone requires vitamin D, a substance that can also stimulate bone calcium release Neither compound is completely active in the absence of the other

A second site of PTH action is on the kidney, where it increases reabsorption of calcium and decreases calcium excretion Since PTH also increases serum calcium, the net effect often is an increase in urinary calcium concentration in spite of increased tubular reabsorption Often, a more marked effect of PTH is increased excretion of phosphate Effects of PTH infusion on renal excretion of phosphate and cyclic AMP are shown in Figure 7 39 The phosphaturic effect of parathyroid hormone may be mediated by an effect on the cellular movements of calcium rather than on phosphate itself (Eisenberg, 1968)

Whether vitamin D has a direct effect on kidney or not is still being debated, but like PTH it causes phosphaturia It also increases plasma phosphate and urinary calcium, but

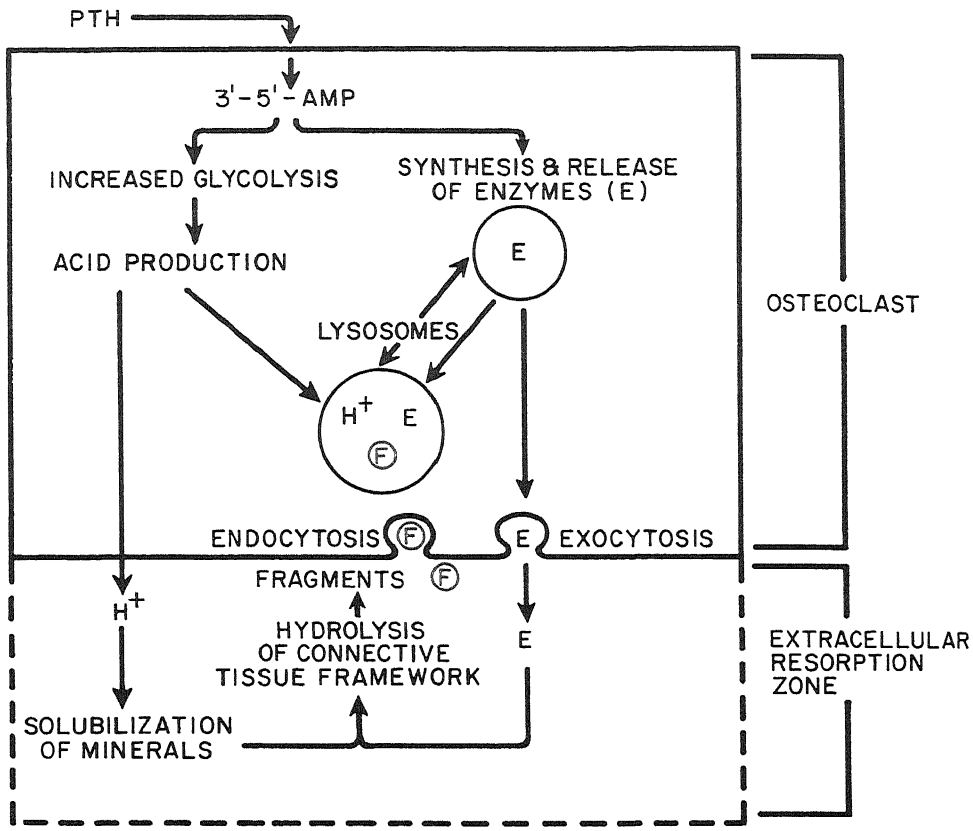


Fig 7 38 Presumed mechanism of PTH-induced bone resorption by osteoclasts (Adapted from Vaes 1968b)

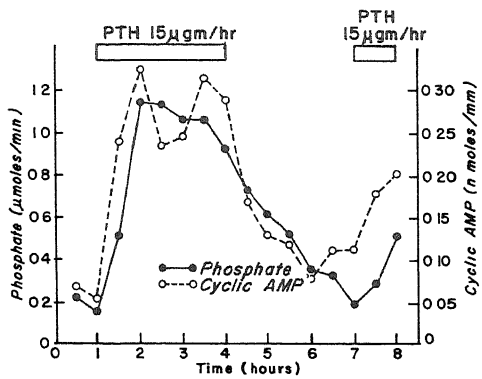


Fig 7 39 Effect of infusion of parathyroid hormone on excretion of phosphate and cyclic-3'-5'-AMP by a parathyroidectomized rat Cyclic AMP appearing in the urine is largely of renal origin and is associated with increased adenyl cyclase activity in cortical tissue In contrast vasopressin increased adenyl cyclase activity in the renal medulla (From Chase and Aurbach 1968)

does not increase calcium reabsorption from the tubule as PTH does

A third site of PTH action is the intestine, where it increases active absorption of calcium, but only in cooperation with vitamin D Some active transport does occur without PTH, but none without the vitamin Phosphate transport by small intestine is also stimulated by PTH and vitamin D

THYROID C CELLS

During the course of studies of parathyroid hormone, several kinds of experiments could not be explained by invoking a calcium-raising hormone alone, but suggested the existence of a calcium-lowering substance as well For example, when thyroparathyroids of dogs were perfused with a high calcium solution, blood calcium fell further and faster than it did in animals that were parathyroidectomized Copp et al (1962) proposed that lowering of serum calcium when the glands were perfused with high calcium was not simply due to shutting off of PTH secretion by negative feedback Munson et al (1968) then found that injection of thyroid extracts decreased serum calcium by 2 mg /100 ml (Fig 7 40)

Since this discovery in the early 1960's, cal-

citonin has been isolated and purified and the structure has been determined Comparative structure of porcine, bovine, salmon, and human calcitonin is shown in Figure 7 41 Mammalian hormones were obtained from extracts of thyroid while salmon calcitonin was extracted from the ultimobranchial body, an outgrowth of the primitive pharynx Each hormone consists of a single chain of 32 amino acids with an intrachain disulfide bridge between positions 1 and 7 and a prolinamide at the carboxyl terminus When tested in mammals, biological activities of porcine, bovine, and human calcitonin are similar, but the salmon hormone is 20 to 40 times more active Only nine positions are occupied by the same amino acid in all 4 molecules Some regularity of structure is apparent, however, since hydrophobic residues occur at almost regular intervals along the molecule Synthesis of the entire molecule and various peptides has shown that all 32 amino acids are necessary for full biological activity (Sieber et al, 1970)

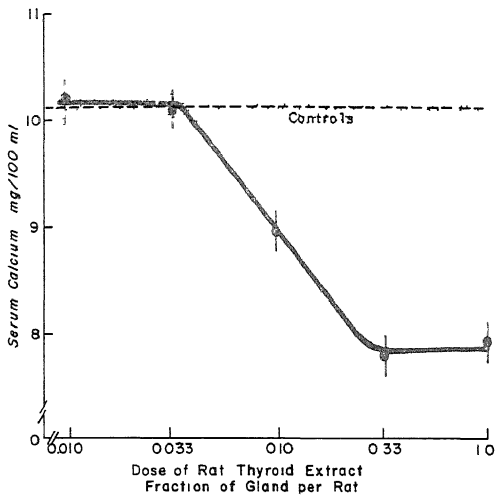


Fig 7 40 Dose response curve illustrating the effect of rat thyroid extract on serum calcium in young male rats 1 hr after subcutaneous infection The horizontal dashed line and shaded band at the top represent the mean and standard error for control rats (From Munson et al 1968)

Calcitonin is synthesized in the parafollicular "C" cells of the thyroid, cells which came originally from embryonic tissue of the ultimobranchial body Immunofluorescence studies have been used to identify calcitonin-producing cells (Fig 7 42) The hormone is stored in the cells in granules measuring 100 to 200 μ When the gland is perfused with hypercalcemic blood, stored hormone is released as indicated by degranulation of the cells Secretion of calcitonin is also stimulated by glucagon and cyclic 3'-5'-AMP, suggesting a role for cyclic-AMP in secretion of this protein (Bell, 1970) Cytological studies indicate that

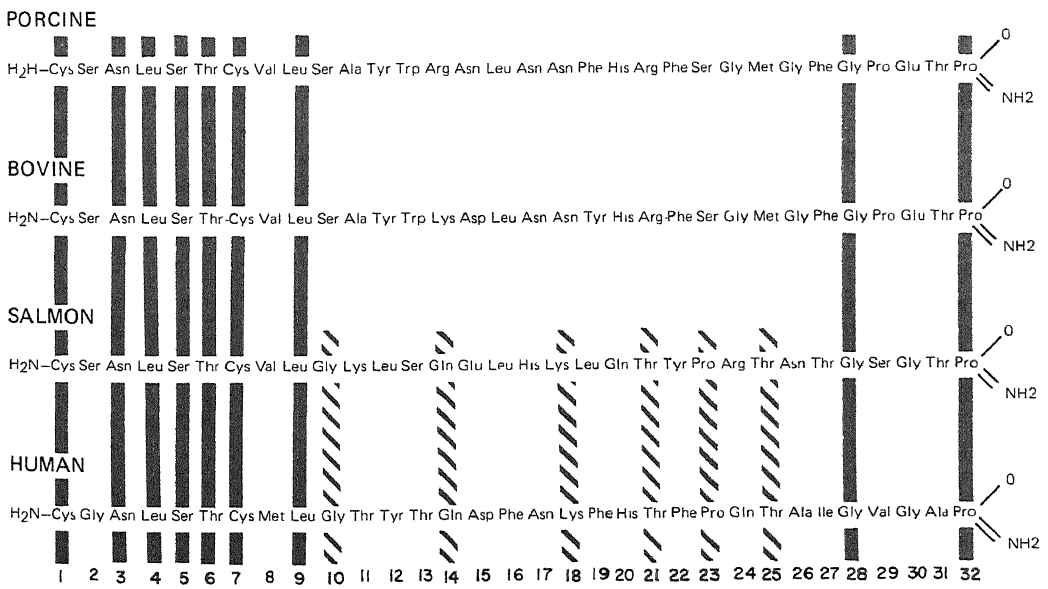


Fig 7 41 Comparison of amino acid sequence of porcine bovine human and salmon calcitonin Solid bars indicate sequence positions homologous between salmon and human hormone stippled bar indicates position where either phenylalanine or tyrosine is found in each of the calcitonins (From Potts et al 1970)

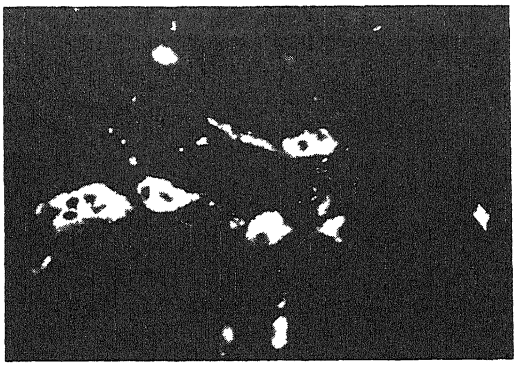


Fig 7 42 Dog thyroid gland The tissue was freeze-dried at −40° C and embedded in paraffin wax Specific immunofluorescence exclusively in C cells was demonstrated using an indirect technique with rabbit antiporcine calcitonin and fluorescence-labelled goat anti-rabbit γ-globulin (From Pearse 1970)

much more calcitonin is stored and ready for release by the "C" cells than PTH by the chief cells of the parathyroid.

Immunoassay of calcitonin has been developed and has shown *1)* that calcitonin is secreted at physiological levels of blood calcium, *2)* that secretion of hormone is raised within a few minutes after blood calcium is raised, and *3)* that the hormone rapidly disappears from the blood The large increase in secretion following perfusion with blood containing high calcium levels confirms the presence of large stores of preformed calcitonin Calcitonin release is proportional to the blood calcium level within 9 to 22 mg /100 ml Inspection of Figure 7 43 shows that PTH and calcitonin are secreted together only within the range of normal plasma calcium levels

ACTIONS OF CALCITONIN

The primary effect of calcitonin is to inhibit bone resorption and calcium release, thereby lowering blood calcium The maximum effect is a decrease of 3 to 4 mg /100 ml of calcium Although not directly competitive to PTH, the effect is opposite. As with PTH, magnitude of the effect on bone is significantly greater in young animals Calcitonin causes a rapid decrease in plasma phosphate as well as calcium The mechanisms of these effects are not understood, but the hormone in some way abolishes the effectiveness of PTH after formation of cyclic AMP. Actually, the bone effects of PTH and calcitonin are somewhat more complicated, since in addition to effects on calcium flux, PTH increases the rate of bone collagen formation, calcitonin inhibits it, and both together will further increase collagen formation (Kalu, et al, 1970) No definite effect of calcitonin has been shown on kidney or gut, and its full effect may be seen even in eviscerated animals

Hypercalcemia is fairly rare under normal circumstances, and this has raised a question about the relative importance of calcitonin The hormone can protect an animal from hypercalcemia after a forced feeding of high amounts of calcium, but recent experiments show that it can also prevent hypercalcemia in a fasted rat eating a normal ion-containing diet (Munson and Gray, 1970) It is likely to counterbalance PTH action in all situations of calcium stress, such as lactation in mammals, starvation, and periods of rapid bone growth

DISORDERS OF CALCIUM HOMEOSTASIS OF C CELL ORIGIN

Calcium imbalances deriving from calcitonin are rare, but it has been proven that medullary carcinomas of the thyroid gland are really tumors of C cells, and these produce large amounts of calcitonin Blood calcium and phosphate may be normal or in some instances low In these cases, PTH levels are elevated Surgery is the only treatment available Calcitonin itself is being tested clinically to correct conditions such as Paget's disease or osteitis deformans and osteoporosis, diseases where bone formation is less than normal Calcitonin extracted from salmon has proven to be many times as effective as that from other species, but use of any protein hormone is complicated

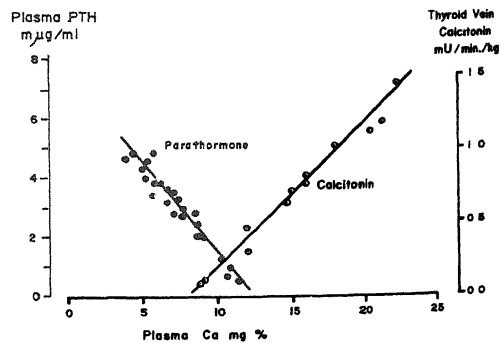


Fig 7 43 Effect of plasma calcium on plasma levels of parathyroid hormone and calcitonin secretion rate (From Copp 1968)

by recipient formation of antibodies to it (Potts, 1970)

DISORDERS OF CALCIUM HOMEOSTASIS OF PARATHYROID ORIGIN

The most common cause of hypoparathyroidism is damage to the glands during thyroid surgery, which is estimated to occur in 1 out of 100 operations However, idiopathic lack of function is found This condition leads to a fall in plasma calcium and a rise in plasma phosphate Urinary calcium increases initially due to lack of reabsorption by the tubules, but falls later as plasma calcium declines Urinary phosphate is low The best known symptom is tetany, which involves muscle spasms and hyperreflexia, general tonicoclonic convulsions, laryngeal spasms, and then death from asphyxia These changes result from a lowering of the threshold for nerve and muscle excitation, so that a nerve may be fired by a slight stimulus anywhere along its length Transmission across synapses is not impaired since the extracellular level of calcium does not get low enough to prevent acetylcholine release

The obvious treatment for hypoparathyroidism would be to raise calcium levels with PTH, but there is little pure hormone available Rather than PTH, very large doses of vitamin D are used along with a high calcium-low phosphate diet

Hyperparathyroidism is usually related to a tumor In this situation, there is no negative feedback and PTH secretion continues even when calcium is high This leads to high blood calcium, low phosphate and increased bone resorption Common findings are bradycardia, ventricular arrhythmias, muscular weakness, urinary stones or soft-tissue calcification, osteoporosis, and ulcers from increased gastric acid secretion Many of the dysfunctions are due to an increased threshold for nerve and muscle depolarization, and precipitation of excess salts in soft tissues Treatment is usually surgery

Secondary hyperparathyroidism can occur from vitamin D deficiency or low calcium intake In these instances, plasma calcium is low, and thus PTH secretion is stimulated Bone resorption will not provide enough ion if absorption continues to be low The continuous stimulus for PTH secretion will lead to hypertrophy of the glands Plasma phosphate may be low, urinary phosphate high, and bone mineralization impaired, leading to ricketts

6 / FUNCTION OF THE ADRENAL GLANDS

The importance of the adrenal glands in maintaining life was first recognized clinically by Addison in 1855 and confirmed experimentally by Brown-Sequard in the following year (Bishop, 1955) The adrenal is, in fact, a compound gland-within-a-gland The outer portion, the cortex, is derived from mesodermal cells which bud from the celomic epithelium It secretes hormonal steroids, as do the ovaries and testes, which are formed from similar mesodermal cells

The developing adrenal cortex is penetrated by chromaffin cells which migrate from the neural crest ectoderm and ultimately form the inner portion of the adrenal gland, the medulla The several endocrine tissues derived from similar neurectodermal cells secrete hormones synthesized from amino acid precursors the polypeptide hormones or, in the case of the adrenal medulla, the catecholamines

The purpose of this chapter is to review very briefly the biosynthesis of the adrenal hormones, the mechanisms controlling their synthesis and release, their general metabolic actions, and the pathophysiology of certain clinical disorders of adrenal function

The Adrenal Cortex

DEVELOPMENT AND HISTOLOGY

The human fetal adrenal is well developed after the first 10 to 12 weeks of gestation The cortex consists of a thin outer rim of small ovoid basophil cells destined to become the definitive adult cortex, and a much wider inner zone of large acidophil cells arranged in anastamosing cords, the fetal cortex The fetal cortex, which comprises 90% of the cortex at term, requires stimulation by fetal pituitary adrenocorticotropic hormone (ACTH) for its development (Benirschke, 1956, Lanman, 1962) The fetal adrenal is larger, in relation to total body weight, than the adult gland and has the ability to synthesize both cortisol and the C_{19} steroids (Bloch and Benirschke, 1959) The fetal cortex begins to involute at about the time of birth and completes this process in 6 to 18 months

The adult human cortex, which differentiates in infancy as the fetal cortex involutes, is divided into three fairly distinct zones the outermost zona glomerulosa, which is concerned primarily with aldosterone synthesis, and the wide middle zona fasciculata and innermost zona reticularis, which together are concerned with cortisol production In some mammals the reticularis is primarily responsible for the synthesis of adrenal androgens It was once thought that the adrenal cortical cells originated in the zona glomerulosa, migrated inward through the fasciculata, and degenerated in the reticularis Although it is true that cell generation appears to be most rapid in the zona glomerulosa (Ford and Young, 1961, Walker and Rennels, 1961), and that a complete adrenal cortex can regenerate from a few cells adherent to the capsule (Penney et al, 1963), there is little evidence that these cells mature and migrate centripetally under normal conditions The ultrastructure of zona glomerulosa cells reveals elongated mitochondria with lamellar cristae, cytoplasm filled with finely vesiculated smooth endoplasmic reticulum, and membrane enclosed lipid droplets The fasciculata cells have round to ovoid mitochondria with vesicular cristae which provide an exceptionally large inner surface area These cells are filled with the abundant smooth endoplasmic reticulum commonly observed in steroid secreting cells, and also contain liposomes The cells of the zona reticularis resemble those of the fasciculata, but have more variable mitochondrial size and appearance (Luse, 1967)

BIOSYNTHESIS OF ADRENAL CORTICOSTEROIDS (Samuels and Uchikawa, 1967 Bransome, 1968)

The adrenal cortex synthesizes three classes of hormones glucocorticoid, mineralocorticoid, and sex hormone The most important steroids secreted by the normal adult human adrenal cortex are the glucocorticoid, cortisol, and the mineralocorticoid, aldosterone In addition, small quantities of estrogens and steroid precursors, including progesterone, are released into the circulation Adrenal androgens, al-

though secreted in large quantities, are of low physiological potency

Glucocorticoid (Cortisol) The hormonal steroids are all synthesized from acetyl coenzyme A with cholesterol as their common intermediate (Caspi et al, 1962) Only a small fraction of the intracellular cholesterol pool is immediately available for steroid synthesis (Hechter et al, 1953, Werbin and Chaikoff, 1961) This small pool appears to be in rapid equilibrium with plasma cholesterol (Borkowski, et al, 1967) The uptake of cholesterol from plasma, furthermore, appears to be stimulated by ACTH (Dexter et al, 1967)

The major reactions in the biosynthesis of adrenal steroids are represented in Figure 7 44 While even this very simplified diagram at first glance appears complex, the biosynthetic process actually represents a series of discrete, rather minor modifications of the cholesterol molecule One needs to understand these pathways in order to understand the pathophysiology of the congenital adrenal hyperplasias The exact sequence in which the reactions occur is not known and may depend upon the relative concentrations of substrates, enzymes, and cofactors The probable major pathways are indicated by the heavy arrows The cleavage of the side chain of cholesterol to form pregnenolone takes place in the mitochondrion Cleavage begins with the formation of 20α-hydroxycholesterol This is the rate-limiting step in steroid synthesis and is accelerated by ACTH in the zona fasciculata and, presumably, by angiotensin II in the zona glomerulosa Conversion to pregnenolone is completed in the mitochondrion by what is referred to as the desmolase reaction (Shimizu et al, 1961) There is an alternate pathway which circumvents pregnenolone by way of 17-hydroxypregnenolone (Shimizu et al, 1965) Pregnenolone now moves out of the mitochondrion The 3β-hydroxyl group of pregnenolone is oxidized to a ketone by 3β-ol dehydrogenase, located on the smooth endoplasmic reticulum This reaction requires nicotinamide adenine dinucleotide (NAD) as the hydrogen acceptor (Beyer and Samuels, 1956) The shift of the double bond is catalyzed by Δ^5-3-ketosteroid isomerase (Ewald et al, 1964), forming progesterone

The remaining reactions in the synthesis of cortisol are all hydroxylations and require as cofactors molecular oxygen, reduced nicotinamide adenine dinucleotide phosphate (NADPH), flavoproteins, the cytochrome system, a nonheme iron protein, and a special cytochrome called cytochrome P-450 because of its sharp absorption peak at 450 mμ after exposure to carbon monoxide An electron is transferred to cytochrome P-450, which in turn reduces molecular oxygen to water and an activated oxygen atom, which is inserted between the steroid ring carbon and its appropriate α or β hydrogen (Harding and Nelson, 1966) Since the zona glomerulosa lacks 17α-hydroxylase, progesterone is converted to 17-hydroxyprogesterone only in the zonae fasciculata and reticularis, where it is then converted to 11-deoxycortisol (Reichstein's "substance S") the final, relatively inactive precursor of cortisol Deoxycortisol then returns from the smooth endoplasmic reticulum to the mitochondrion for the final 11β-hydroxylation to cortisol ("substance F") The normal rate of cortisol synthesis is 15 to 30 mg /day

An intriguing historical facet of the study of glucocorticoid biosynthesis has been the study of the role of adrenal ascorbic acid in this process Ascorbic acid is present in very high concentrations in the normal adrenal gland (Szent-Gyorgyi, 1928) Adrenal ascorbic acid decreases as steroidogenesis is stimulated by ACTH Despite a great deal of interest, speculation, and investigation, the exact relationship of ascorbic acid to steroid synthesis remains a mystery

Mineralocorticoid (Aldosterone). Since 17α-hydroxylase is absent in the zona glomerulosa, progesterone is converted on the smooth endoplasmic reticulum of the glomerulosa first to 11-deoxycorticosterone (DOC) by 21-hydroxylation (Fig 7 44) and then, by 11β-hydroxylation, within the mitochondrion, to corticosterone, "substance B " The mitochondrial enzyme, 18-hydroxylase, catalyzes its conversion to 18-hydroxycorticosterone 18-hydroxycorticosterone is dehydrogenated, by an enzyme found only in the mitochondria of the zona glomerulosa, to aldosterone (Wettstein et al, 1955) It is secreted at a rate of 50 to 200 μg /day

Androgens and Estrogens The side chain of 17α-hydroxypregnenolone is cleaved at the 17-20 carbon bond (Fig 7 44) to form dehydroisoandrosterone (DHA) which, with its sulfate, is the major precursor of urinary 17-ketosteroids Minor quantities of 17-hydroxypro-

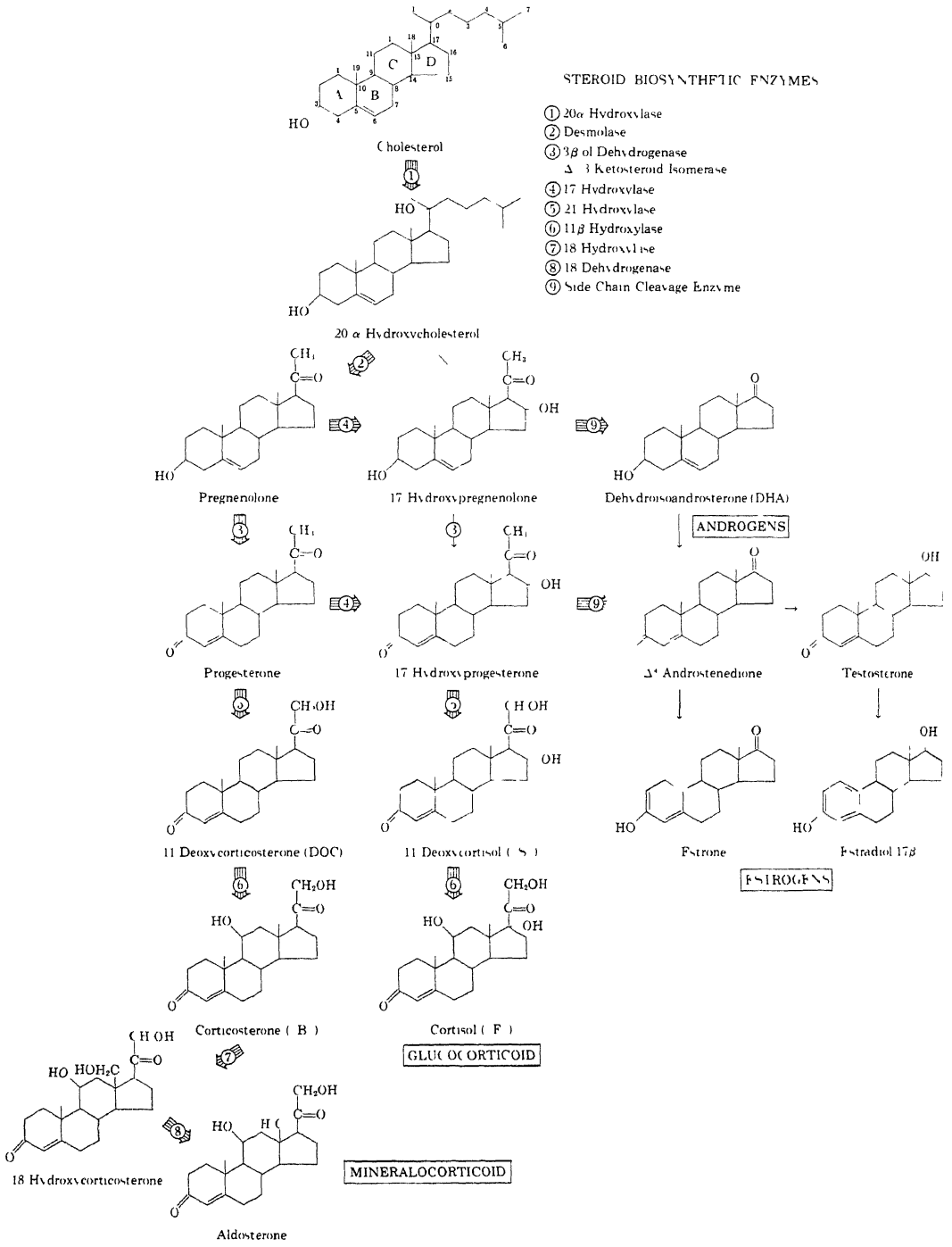


Fig 7 44 A simplified scheme showing the pathways in adrenocortical steroid biosynthesis after formation of cholesterol The probable major reactions are indicated by heavy arrows

gesterone are converted by a similar reaction to another 17-ketosteroid, Δ^4-androstenedione, which can be further converted, by reduction of the 17-ketone, to testosterone Estrogens apparently can be produced by adrenal tissue, as evidenced by estradiol and estrone production, in vitro by feminizing adrenal tumor tissue (Dorfman et al , 1965) The

amounts normally produced, if any, are very small, however, and the biosynthetic pathways are not clear They may involve 19-oxidation of testosterone, followed by 19-demethylation and aromatization of the A ring to yield estradiol-17β A similar series of reactions might produce estrone from Δ^4-androstenedione More important, DHA sulfate, secreted by the adrenal, can be converted to estrogens by the ovaries (MacDonald et al , 1967) Accurate secretory rates for adrenal androgens and estrogens are difficult to establish because of gonadal secretion and peripheral interconversion of these steroids

METABOLISM OF ADRENAL CORTICOSTEROIDS

Glucocorticoid (Cortisol). Cortisol is reversibly oxidized to cortisone ("substance E") in the tissues, predominantly the liver Cortisone is biologically inert and has a half-life in plasma less than half that of cortisol Administered cortisone has about two-thirds the biological potency of cortisol only by virtue of its bioconversion to cortisol

In plasma, about 2% of cortisol normally circulates in the free state, a small fraction is nonspecifically bound to serum albumin, and greater than 90% is specifically bound to corticosteroid-binding α_2-globulin (CBG) (Daughaday, 1967) The normal CBG concentration is about 3 mg /100 ml , but is increased markedly during pregnancy or estrogen administration The normal CBG binding capacity is approximately 20 μg cortisol/100 ml Normal plasma cortisol concentrations are less than 25 μg /100 ml One molecule of steroid is bound by each molecule of CBG, the strength of binding depending upon the molecular configuration of the steroid The cortisol bound to CBG or albumin is metabolically inactive, but is in rapid equilibrium with free cortisol Thus, the binding proteins act as a reservoir for cortisol, releasing it when free cortisol is removed from the circulating plasma After CBG is saturated with cortisol, however, relatively small increases in plasma cortisol result in a disproportionate increase in the free, metabolically active cortisol fraction The high plasma cortisol levels seen during pregnancy or estrogen administration are secondary to increased CBG The metabolically active plasma free cortisol levels are not proportionately elevated

The major pathway of cortisol degradation involves reduction and conjugation reactions Reductases in the liver perform the major irreversible step in the metabolism of cortisol and cortisone reducing the 4-5 double bond of the A ring The reactions require NADPH as cofactor, the predominant products are the 5β, rather than the 5α-metabolites (Fig 7 45) Further reduction of tetrahydrocortisol (THF) and tetrahydrocortisone (THE) at the C_{20} carbonyl yields the cortols and cortolones By far the major portion of the metabolites of cortisol and cortisone are excreted in the urine as conjugates of glucuronic acid Esterification to the glucuronide is catalyzed in the liver by glucuronyl transferase, which transfers glucuronic acid from uridine diphosphate (UDP)-glucuronic acid to a hydroxyl radical, usually the 3β-hydroxyl of tetrahydrocortisol (Fig 7 45) Cortisol is cleared from plasma with a disappearance half-time of 70 to 80 min (Peterson, 1959)

Mineralocorticoid (Aldosterone) Aldosterone is weakly bound to plasma proteins, perhaps to CBG Under normal conditions, aldosterone circulates in the hemiacetal form, which is more stable than the aldehyde configuration (Fig 7 46) The major metabolic pathway for aldosterone degradation is shown Less than 1% of aldosterone is excreted in the urine unaltered The major portion of aldosterone, however, is excreted as tetrahydroaldosterone glucuronide As is the case with the glucocorticoids, the product is formed by the action of hepatic glucuronyl transferase and is conjugated through the 3β-hydroxyl of tetrahydroaldosterone The normal half-life of aldosterone in plasma is approximately 30 min (Coppage et al , 1962)

ASSAY OF ADRENAL CORTICOSTEROIDS

Plasma Corticosteroids. There are four clinically useful methods for measuring corticosteroids in plasma

1) Colorimetric When 17,21-dihydroxy-20-ketosteroids such as cortisol, deoxycortisol, and cortisone react with acid phenylhydrazine, they form yellow phenylhydrazones, the color intensity of which are proportional to the quantities of steroids present, this is the Porter-Silber reaction (Porter and Silber, 1950) Since deoxycortisol and cortisone are normally present only in minor quantities in human plasma, cortisol is the only 17-hydrox-

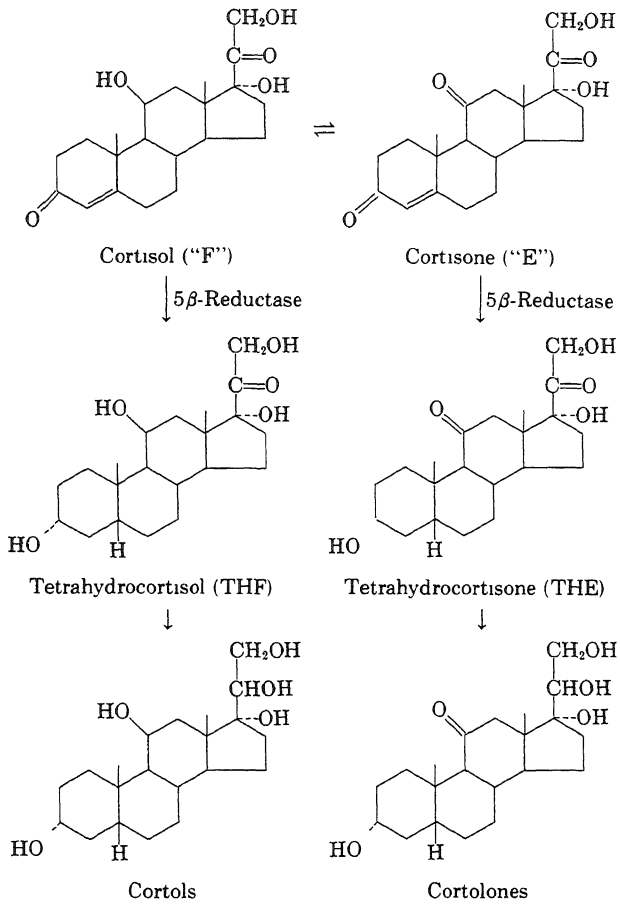


Fig 7 45 The major metabolites of cortisol and cortisone

ycorticosteroid (17-OHCS) of any importance Since many drugs are chromogenic and interfere in this assay, however, they must be excluded by extracting cortisol into lipid solvents, and washing the extract with alkali

2) *Fluorometric* Steroids extracted from plasma first into dichloromethane and then into ethanolic sulfuric acid can be quantitated by their fluorescence (DeMoor et al , 1960, Mattingly, 1962) The structural requirement is the Δ^4-11β,21-dihydroxy-3,20-dione configuration Thus, both cortisol and corticosterone are measured, deoxycortisol is not Since corticosterone is present in only small quantities in human plasma, plasma fluorogenic steroid (11-OHCS) is equivalent to cortisol Very few drugs interfere in this assay An exception is the aldosterone antagonist, spironolactone

3) *Competitive Protein-Binding* (Murphy, 1969) A derivative of the radioimmunoassay technique (Berson and Yalow, 1967, 1968), this assay utilizes CBG, rather than an antibody, to bind the steroid hormone Tracer amounts of labeled steroid are added to a very small concentration of CBG and incubated The percent of labeled steroid bound to CBG is determined As increasing quantities of unlabeled reference standard steroid are added, they compete with the labeled steroid for CBG-binding sites, "displacing" label from the CBG A standard curve can be constructed, plotting the decrease in per cent binding of labeled tracer against increasing concentration of unlabeled standard The concentration of steroid in an unknown can be determined by comparing its ability to displace tracer with that of the reference standard

4) *Radioimmunoassay.* Since aldosterone binds so slightly to CBG or other plasma proteins, it is difficult to measure by the competitive protein-binding method However, steroid

molecules which are not themselves antigenic can be covalently conjugated to larger molecules, such as serum albumin When the conjugate is injected into a recipient animal, antibodies may be formed to the steroid, which acts as a haptenic group These antibodies may bind the free steroid itself, and thus be useful for radioimmunoassay, as has been the case in the assay of plasma aldosterone (Mayes et al , 1970, Bayard et al , 1970) Plasma concentrations in normal subjects on unrestricted sodium diets are between 1 to 16 ng /100 ml of plasma or less

Urinary Corticosteroids. Free cortisol is excreted in the urine in very small quantities —40 to 200 μg /24 hr. Since the great majority of urinary steroids are water-soluble glucuronide conjugates, they must be hydrolyzed with β-glucuronidase in order to extract the free steroids into a lipid solvent

1) 17-Ketosteroids (17-KS) The oldest of the urinary steroid assays, this method utilizes the reaction of the steroid with metadinitrobenzene (Zimmerman's reagent) to form a reddish-purple color, the intensity of which is proprotional to the quantity of steroid present (Zimmerman, 1935, Callow et al , 1938) The major precursor of urinary 17-ketosteroids is dehydroisoandrosterone, the production of which does not always correlate well with cortisol secretion Thus, urinary 17-KS are not a reliable indicator of adrenal glucocorticoid production, but do reflect the secretion of the adrenal androgens except testosterone, which is not a 17-ketosteroid Normal values are usually given as 5 to 15 mg /24 hr in the female, and 10 to 22 mg /24 hr in the male, children less than 9 excrete less than 3 mg /day, but this increases to adult values during puberty

2) 17-Hydroxycorticosteroids (17-OHCS) The Porter-Silber reaction may be used to measure urinary 17,21-dihydroxy-20-ketosteroids Approximately 30 to 40% of secreted cortisol appears in the urine as THF and THE, tetrahydrodeoxycortisol (THS) is normally present in only small quantities While normal values are generally given as 3 to 12 mg /24 hr , cortisol secretion is a function of body size Since urinary creatinine excretion is proportional to lean body mass and generally is very constant from day to day, normal urinary 17-OHCS can better be expressed as 3 to 7 mg /g of creatinine In obese persons, 17-OHCS excretion is greater than one might expect on the basis of lean body mass and normally amounts to less than 0 12 mg /kg /24 hr In children, normal excretion can conveniently be related to body surface area 1 7 to 4 0 mg /M^2/24 hr

3) 17-Ketogenic Steroids (17-KGS) In addition to the 17-OHCS, certain other steroids,

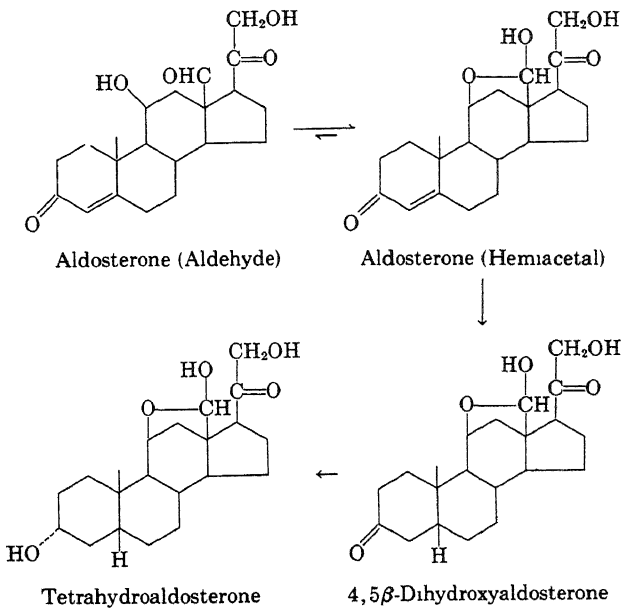


Fig 7 46 The major metabolites of aldosterone Aldosterone is shown both in the aldehyde configuration and the more stable hemiacetal form

such as pregnanetriol, can be oxidized with sodium bismuthate or sodium metaperiodate to 17-KS and measured by the Zimmerman reaction (Norymberski, 1961, Rutherford and Nelson, 1963) They are a less specific indicator of cortisol secretion than the 17-OHCS, but are nevertheless determined in some laboratories The preformed 17-KS must either be determined separately and subtracted from the total 17-KGS, or must first be reduced to a form nonreactive with the Zimmerman reagent Normal values are given as 8 to 20 mg /24 hr

4) Aldosterone Excretion Rate Aldosterone is present in urine mainly as tetrahydroaldosterone glucuronide Thus, it must first be hydrolyzed with β-glucuronidase before extraction into lipid solvents It may then be measured by its fluorescence If the determination is performed on an aliquot of a 24-hr urine collection, the aldosterone excretion rate can be calculated The normal range is 2 to 16 μg /24 hr (Mattox and Lewbart, 1959) For a variety of reasons, however, this method is not a very satisfactory index of aldosterone secretion

5) Aldosterone Secretion Rate A much more definitive procedure is the determination of the aldosterone secretion rate (ASR) by the double isotopic dilution method In this technique, a single tracer dose of tritium-labeled aldosterone is administered to the patient intravenously and the urine is collected for the following 24 hr The administered radioactive aldosterone mixes with that secreted by the adrenal, and both are assumed to be metabolized and excreted by the kidney in an identical manner Approximately 80 to 90% of the administered counts are recovered The aldosterone metabolites are isolated by paper chromatography and are then labeled by acetylation with ^{14}C-acetic anhydride The fraction of the total (^{14}C-labeled) aldosterone accounted for by the administered (^{3}H-labeled) aldosterone will be inversely proportional to the size of the endogenous pool in which it was diluted, the dilution factor is, thus, a measure of the amount of aldosterone secreted, metabolized, and excreted during the 24-hr period Knowing the dilution factor and the exact amount of ^{3}H-labeled aldosterone administered, an accurate estimate of endogenous aldosterone secretion rate can be made Under normal basal conditions on a regular diet, the ASR is 50 to 200 μg /24 hr

PHYSIOLOGICAL EFFECTS OF ADRENAL CORTICOSTEROIDS

There is as yet no unifying concept to explain all of the complex actions of cortisol, which affect virtually every tissue in the body Clinically, cortisol's actions might almost be described as "permissive" (Sayers, 1950), without cortisol secretion, a multitude of seemingly diverse physiological processes fail to function optimally

The term, glucocorticoid, refers to the sum of a wide range of influences on organic metabolism (carbohydrate, fat, and protein) and includes, among others, effects on the immune and inflammatory responses, wound healing, vascular permeability, and muscle integrity Mineralocorticoid effects, on the other hand, are rather well-defined as the sodium-retaining and potassium-wasting effects on the distal renal tubule, the salivary glands, the sweat glands, and the gastrointestinal tract Glucocorticoid and mineralocorticoid are artificial terms, since most natural and synthetic corticosteroids possess at least some of each activity For all practical purposes, however, aldosterone may be considered as a pure mineralocorticoid and dexamethasone, a steroid that is chemically synthetized, as a pure glucocorticoid (Table 7 5) While cortisol has a significant sodium-retaining activity, its effect at normal plasma cortisol concentrations is relatively minor compared to that of aldosterone

General Metabolic Effects of Cortisol (Ashmore and Morgan, 1967) The overall pattern of cortisol action might be viewed as promoting the conversion of protein to carbohydrate and the storage of carbohydrate in the form of glycogen

1) Protein Metabolism Adrenalectomized rats have decreased urea nitrogen excretion, blood glucose concentration, and liver glycogen storage, indicating decreased protein catabolism and decreased gluconeogenesis (Long et al , 1940) These aberrations are corrected with administration of adrenal steroids Glucocorticoids mobilize amino acids from skeletal muscle protein (Bondy et al , 1949, Engel, 1950, Kaplan and Shimizu, 1963) and exert a profound influence on hepatic metabolism Stimulation of the uptake of amino acids from plasma by the liver (Riggs, 1964) and induction of the synthesis of a number of hepatic enzymes, among them the transaminases, pre-

Table 7 5
Relative potencies of various natural and synthetic steroids

Steroid	Glucocorticoid Potency (Cortisol = 1)	Mineralocorticoid Potency (Aldosterone = 1)	Glucocorticoid/ Mineralocorticoid	Daily Secretion Rate or Maintenance Dose
Aldosterone	0 1	1 00	0 005	0 05-0 2 mg
Fludrocortisone	10 0	0 21	2 4	0 1 mg
Deoxycorticosterone	~0	0 05	~0	2 5 mg
Cortisone	0 7	0 0012	27	25-37 5 mg
Cortisol	1 0	0 0015	33	15-30 mg
Prednisone	4 0	0 0020	100	5-7 5 mg
Dexamethasone	30 0	~0	∞	0 5-0 75 mg

pare amino acids for incorporation into carbohydrate (Rosen et al , 1963) Since activities of enzymes in the urea cycle are also increased, conversion of the amino nitrogen to urea is facilitated (McLean and Gurney, 1963) The glucocorticoids, therefore, result in what has been called a "translocation" of protein skeletal muscle protein is depleted, but the synthesis of proteins by the liver and other visceral organs is enhanced by steroid administration (Roberts, 1953, Silber and Porter, 1953) There is an overall negative nitrogen balance

2) *Carbohydrate Metabolism* The action of cortisol is required to maintain normal hepatic production of glucose Cortisol appears to stimulate glycogen synthesis by increasing the conversion of pyruvate to glycogen and by inhibiting the release of glucose from the hepatic cell The net effect of glucocorticoid action on carbohydrate metabolism is increased synthesis and storage and decreased utilization, with a resulting hyperglycemic tendency

3) *Lipid Metabolism* The role of cortisol in lipid metabolism is not well understood Total body fat is increased by glucocorticoid administration (Kyle et al , 1956), but fatty acid synthesis in the liver is inhibited by cortisol (Ashmore et al , 1956) This effect is not observed in adipose tissue, however, where synthesis of fatty acids appears to be regulated primarily by the rate of glucose utilization (Jeanrenaud and Renold, 1960) Cortisol does promote the mobilization of fatty acids from peripheral adipose tissues and plays a permissive role in the lipolytic actions of hormones such as thyroxine and the catecholamines, a minimal amount of cortisol must be present for these hormones to exert their lipolytic effects (Goodman and Knobil, 1961)

4) *Nucleic Acid Metabolism.* Cortisol inhibits the synthesis of ribonucleic acid (RNA) and deoxyribonucleic acid (DNA) in lymphoid and connective tissues, thus exerting a catabolic effect, but stimulates RNA synthesis in the liver and other visceral organs The effects of cortisol on the enzyme systems in the liver, like the effects of other steroid hormones on other tissues (O'Malley, 1969), result from stimulation of DNA transcription and the formation of specific messenger RNA species, which in turn regulate the synthesis of specific proteins, including enzymes (Fig 7 47)

Cortisol Effects on Organ Systems. It is easier to define the effects of cortisol deficiency or excess, which are manifested in every organ system, than to define the role of glucocorticoids in normal physiology Some of the effects of cortisol on the various organ systems are listed here

1) *Central Nervous System* Glucocorticoids seem to have a stimulatory effect on brain excitability, independent of electrolyte effects (Woodbury, 1958), and alter electroencephalographic patterns Insomnia, euphoria, irritability and psychotic depression or schizophrenia are associated with hypercortisolemia, inability to concentrate or to perform mathematical calculations and mental fatigue are manifestation of hypocortisolism The mechanisms by which glucocorticoids exert these effects on the central nervous system are unknown

2) *Musculoskeletal System* The catabolic effects of cortisol on skeletal muscle are manifested by muscle atrophy and objective muscle weakness The increased efflux of potassium and amino acids from skeletal muscle in the early morning hours may be a manifestation of the protein catabolic effect of cortisol, which reaches its highest plasma concentration at

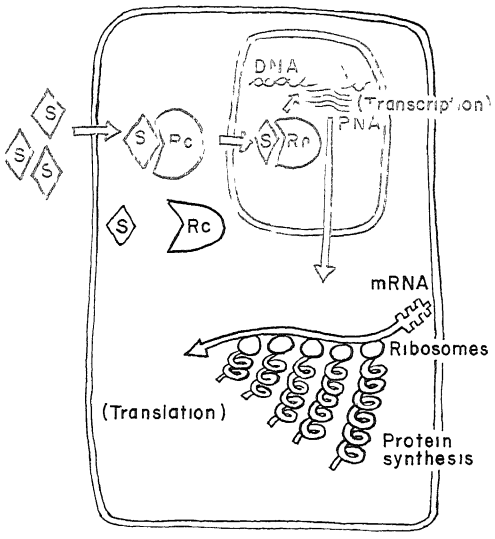


Fig 7 47 A schematic representation of the mechanism of action of a steroid hormone on its target tissue cell The steroid (*S*) diffuses freely into the cell where it is bound by a cytoplasmic receptor protein (*Rc*) The steroid-receptor complex is carried into the nucleus where the receptor may be modified (*Rn*) The complex causes DNA transcription and the synthesis of RNA species Messenger RNA (*mRNA*) directs the synthesis of new proteins on the polyribosomes (translation) (From O Malley 1971)

this time of day (Rabinowitz, 1968) Cortisol has a marked influence on connective tissue integrity In conditions of cortisol excess, peripheral subcutaneous tissue is decreased, the skin becomes thin and friable, and wound healing is delayed Cutaneous striae develop where the skin is stretched These effects may be manifestations of decreased hyaluronic acid and collagen production and inhibition of the growth and function of fibroblasts (Dougherty, 1961, Holden and Adams, 1957) Cortisol inhibits bone formation by decreasing formation of bone osteoid and cartilaginous matrix (Bernick and Ershoff, 1963) Steroids have a hypocalcemic effect, due at least in part to their antagonism to the action of vitamin D in promoting calcium absorption from the gastrointestinal tract

3) Cardiovascular System The maintenance of intravascular volume by sodium retention is mainly the responsibility of aldosterone Cortisol, by virtue of its much greater daily secretory rate, does exert a significant mineralocorticoid effect, however Cortisol appears to potentiate the peripheral arteriolar response to vasoconstrictors It increases cardiac output when given in pharmacologic doses At physiological levels, glucocorticoids decrease the permeability of the vascular endothelium At supraphysiological concentrations, cortisol has a deleterious effect on small vessel integrity

4) Respiratory System Cortisol has no apparent physiological effect on the lungs, although there is evidence that glucocorticoids may be important in the maturation of the fetal lung and its ability to secrete surfactant

5) Gastrointestinal System The antagonism of vitamin D action by cortisol, and cortisol induction of hepatic enzymes have been mentioned While supraphysiological doses of steroid inhibit gastric mucus secretion, there is no conclusive clinical or experimental evidence to support the widely accepted belief that high levels of circulating glucocorticoids can produce ulceration of normal gastric mucosa

6) Genitourinary System Menstrual abnormalities, usually oligomenorrhea, are common in women with hypercortisolism, as is downy hirsutism of the face and trunk More marked hirsutism should suggest the effects of more potent androgens than cortisol

Cortisol acts to increase renal plasma flow, glomerular filtration rate, and free water clearance (Kleeman et al , 1958), which tend to produce a sodium diuresis The mineralocorticoid effect of cortisol, however, increases the reabsorption of sodium in the distal renal tubule In addition, cortisol may decrease the permeability of the distal tubular epithelium to water (Kleeman et al , 1964) Thus, the net effect of cortisol on the kidney is maintenance of the ability to excrete a water load

7) Hematopietic and Lymphoid Systems Glucocorticoids appear to have a permissive role in erythropoiesis, polycythemia is often seen in hypercortisolism, and anemia in adrenal insufficiency Glucocorticoids have a dramatic catabolic effect on lymphoid tissue, manifested by lymphoid and thymic atrophy and lymphopenia in hypercortisolism and by lymphoid hyperplasia in adrenal insufficiency Circulating eosinophils fall rapidly after glucocorticoid administration due to increased destruction, this phenomenon was once used as an index in in vivo adrenocortical activity (Nelson, 1953) As the lymphocytes and eosinophils disappear from the circulating blood, the neutrophil polymorphonuclear leukocytes increase, probably because cortisol impairs

their emigration from the vascular space into the surrounding tissues (Boggs et al , 1965)

8) Effects on Inflammation and Infection The glucocorticoids are involved in almost every phase of the inflammatory response Under conditions of cortisol excess, here is interference with the febrile response, impaired diapedesis of granulocytes into the area of injury, impaired migration of phagocytic cells, decreased ability to localize infection because of impaired connective tissue integrity, impaired bacteriocidal activity of phagocytes, decreased phagocytic activity of the reticuloendothelial system, decreased humoral antibody synthesis, hypogammaglobulinemia, eosinopenia, lymphopenia, thymic atrophy, and stabilization of lysosomal membranes, thereby preventing the release of proteolytic enzymes and kinins which may be important in the inflammatory response (Grant, 1967) In the hypoadrenal patient, on the other hand, infection may elicit an exaggerated febrile and inflammatory response, but the patient suffers from an ill-defined inability to respond to "stress " In normal individuals, plasma cortisol concentrations increase under "stressful" conditions, inability to mount this response may result in adrenal crisis and vascular collapse

General Metabolic Effects of Aldosterone (Mulrow, 1969) The most important biological action of aldosterone is regulation of sodium and potassium excretion by the kidney

1) Direct Effects on the Kidney Aldosterone increases distal tubular sodium resorption and potassium excretion It also enhances wasting of hydrogen, magnesium, and ammonium ions The mechanism by which aldosterone exerts these effects is analagous to that already described for cortisol (Fig 7 48) There is again a specific intracellular receptor protein for the steroid (Fanestil and Edelman, 1966) The aldosterone-receptor complex migrates to the nucleus, where it stimulates DNA-dependent RNA synthesis (Castles and Williamson, 1967) and subsequent protein synthesis The exact nature of the protein synthesized is not known It may be an enzyme which results in generation of adenosine triphosphate (ATP) to provide the energy required for the sodium transport pump (Fanestil et al , 1967), which is shown diagrammatically in Figure 7 48, or a permease which directly facilitates sodium transport across the

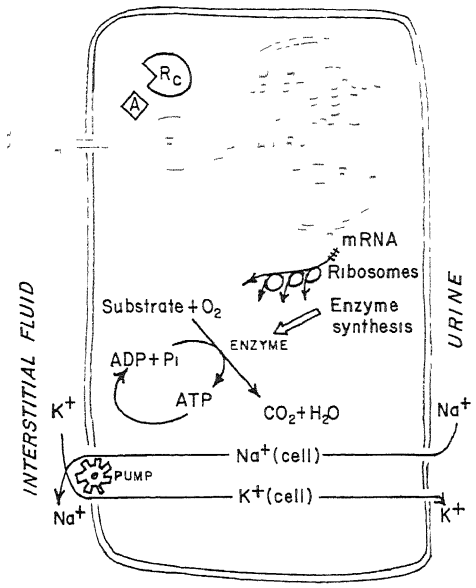


Fig 7 48 A schematic drawing of the mechanism of action of aldosterone (*A*) on the distal renal tubule epithelial cell (From Edelman et al 1964)

cell membrane (Sharp and Leaf, 1966) The effects of aldosterone are facilitated by glucocorticoids but are inhibited by certain cortisol precursors, such as progesterone and 17-hydroxyprogesterone This inhibition is thought to be on the basis of competition for intracellular-binding sites The excretion of potassium normally is coupled directly to sodium resorption Reduced sodium intake or enhanced proximal tubular resorption of sodium, either of which results in less sodium being presented to the distal tubule, prevents the kaliuresis caused by aldosterone (Finn and Welt, 1963) Aldosterone does not affect the excretion of free water by the kidney

2) Indirect Effects on the Kidney The direct sodium-retaining action of aldosterone causes expansion of the plasma volume As a result, renal plasma flow and glomerular filtration increase, renin production decreases After prolonged plasma volume expansion due to sodium retention, the distal tubule exhibits an "escape" phenomenon, it no longer retains sodium in response to aldosterone stimulation, yet it continues to waste potassium (August et al , 1958)

3) Other Metabolic Effects of Aldosterone The sodium-retaining, potassium-wasting effects of aldosterone are expressed in the gastrointestinal tract, salivary glands, and sweat

glands as well as in the renal tubule The difference is that the latent period between administration and observed effect is much longer than the 30 to 120 min delay observed in the kidney and requires de novo protein synthesis Furthermore, these tissues do not "escape" after prolonged administration of mineralocorticoid Aldosterone excess causes hypertension in man, salt deprivation will prevent the hypertension The exact mechanism by which the hypertension is caused is not known It has been suggested that the lumen of the peripheral arteriole is partially occluded by swelling of the arteriolar lining cells, which have retained sodium and water (Tobian, 1960) This partial occlusion would increase peripheral vascular resistance, causing hypertension

REGULATION OF GLUCOCORTICOID SYNTHESIS

The synthesis and secretion of adrenal glucocorticoids are dependent upon stimulation by adrenocorticotropic hormone (ACTH) (Fig 7 49)

Biological Activity. ACTH stimulates the synthesis of cortisol and 17-ketosteroids by the adrenal cortex At physiological concentrations, it has little or no effect on aldosterone secretion ACTH is also necessary to maintain adrenocortical size, structure, and vascularity, growth hormone also plays a minor role in maintaining size In the hypophysectomized animal, the adrenals very rapidly become unresponsive to the acute steroidogenic stimulus of ACTH and eventually become atrophic While secretion of cortisol and adrenal 17-ketosteroids falls to subnormal levels, the secretion of aldosterone remains nearly normal This is because the structure and function of the zona glomerulosa depends, not upon ACTH, but upon the renin-angiotensin system

Actions of ACTH are mediated by cyclic adenosine monophosphate (cyclic AMP) ACTH binds to specific receptor sites on the surface of the adrenocortical cell and activates the associated, membrane-bound adenyl cyclase This promotes the conversion of ATP to cyclic AMP The formation of cyclic AMP is proportional to the amount of ACTH bound and precedes detectable steroidogenesis (Grahame-Smith et al , 1967) There is an intracellular receptor or binding protein for adrenal cyclic AMP, the function of which is not yet known (Gill and Garren, 1969) It is known, however, that ACTH and cyclic AMP stimulate new protein synthesis and that this new protein is required for increased steroidogenesis

This labile protein apparently stimulates the conversion of cholesterol to 20α-hydroxycholesterol in the mitochondrion, either by enhancing the transport of cholesterol from the

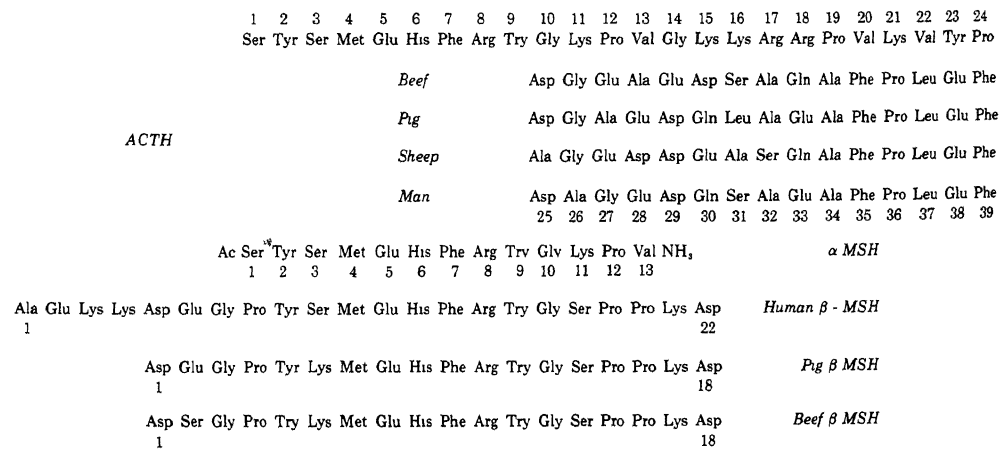


Fig 7 49 The amino acid sequences of representative ACTH s and MSH s The sequence of amino acids 1–24 in ACTH is identical in all four species and possesses full steroidogenic activity (Ney et al , 1965) Species specificity resides in amino acids 25–33 of the ACTH molecule The structure of α-MSH is identical to the first 13 amino acids of ACTH with acetylation of the N-terminal serine and amidation of the C-terminal valine The β-MSH s of three species are shown The dotted lines enclose the sequence of seven amino acids shared by α-MSH the β-MSH s and the ACTH s

liposomes to the mitochondria or by direct stimulation of the hydroxylation reaction

ACTH and the melanocyte-stimulating hormones (MSH's) share certain features They are secreted by the pituitary, they have a common core of seven amino acids (Fig 7 49), and they each have both steroidogenic and melanocyte-stimulating activities One of the extra-adrenal effects of ACTH is the dispersion of melanin pigment in melanocytes, which causes darkening of the skin ACTH has about $^1/_{100}$th the MSH potency of α-MSH The MSH's on the other hand, all have weak intrinsic steroidogenic activity Furthermore, it appears that in man, in whom β-MSH is the principal pigmentary hormone, that ACTH and β-MSH secretion parallel each other, suggesting that they may be synthesized and secreted by a common mechanism within a single cell (Abe et al , 1969a)

Regulation of ACTH Secretion (Liddle et al , 1962) Three factors are of major importance in regulating ACTH secretion First, there is homeostatic negative feedback inhibition of ACTH secretion by circulating cortisol (Fig 7 50) Decreased plasma cortisol concentrations lead to increased pituitary ACTH release, increased plasma cortisol levels inhibit further ACTH secretion This inhibition may act at the hypothalamic level by inhibiting the release of corticotropin-releasing factor (CRF) into the hypothalamic-hypophyseal portal blood vessels, it certainly appears to act at the level of the pituitary, blocking the actions of releasing factors (Fleischer and Vale, 1968) Second, plasma cortisol concentrations show a circadian (circa, about, dies, a day) rhythm, with high levels (15 to 25 μg /100 ml) in the early morning and low levels (less than 7 μg /100 ml) late in the evening (Fig 7 51) This rhythm is the result of a circadian rhythm in ACTH secretion (Ney et al , 1968) and appears to be synchronized by the daily environmental shift from darkness to light (Orth and Island, 1969) There is a constantly changing "set-point" for the concentration of plasma cortisol required to inhibit ACTH secretion during the course of the day, high levels being required in the early morning hours and low levels in the late afternoon and evening Whether or not this is mediated by CRF secretion is not known The third important factor governing ACTH secretion is "stress " Thus, plasma ACTH levels are increased during fever, laparotomy, or acute severe depression, resulting in increased secretion of cortisol While some of the influence of stress may be mediated via hypothalamic CRF release, it is of interest that section of the pituitary stalk not always, and seldom permanently, abolishes stress-induced ACTH release

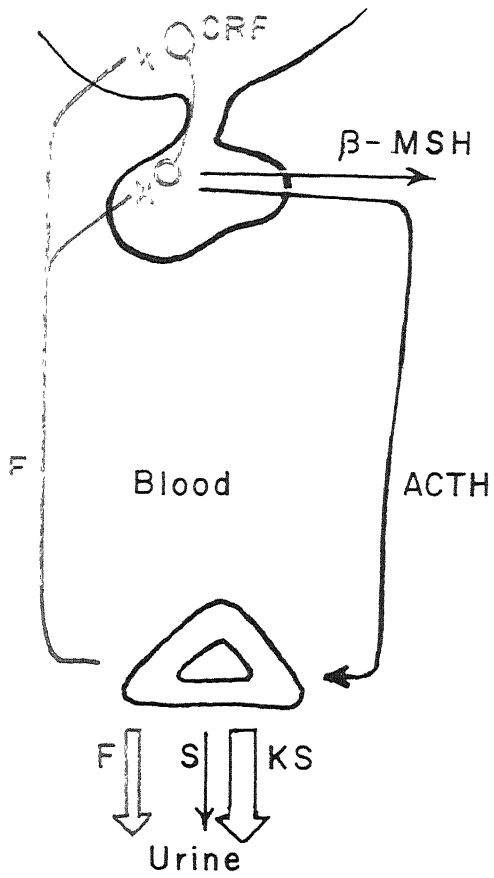


Fig 7 50 A diagram of the normal hypothalamic-pituitary-adrenal axis Corticotropin releasing factor (*CRF*) is released by the hypothalamus into the portal circulation and causes ACTH and *β MSH* release ACTH stimulates adrenal synthesis of cortisol (*F*) Cortisol deoxycortisol (*S*) and 17-ketosteroid (*KS*) metabolites appear in the urine Cortisol exerts a negative feedback inhibition on ACTH secretion and perhaps on CRF secretion Thus decreasing plasma cortisol leads to increasing plasma ACTH Increased plasma ACTH stimulates increased adrenal cortisol production and increased plasma cortisol inhibits further ACTH secretion

Measurement of ACTH. There are currently three basic methods by which ACTH may be assayed The oldest, and in many ways the most reliable, is the bioassay (Lipscomb and Nelson, 1962) Its greatest advantage is specificity for the biologically active ACTH

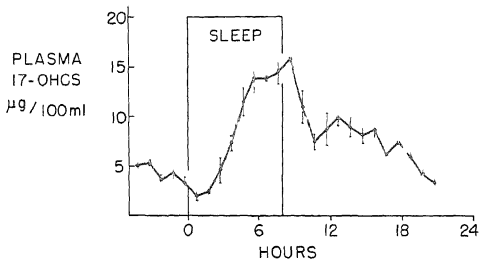


Fig 7 51 Plasma cortisol (17-OHCS) concentrations in eight normal subjects confined to a metabolic ward They slept regularly from 10 PM until 6 AM for several days before and during the period when blood was sampled hourly via indwelling venous catheters The mean 17-OHCS and the standard error of the means are shown

molecule Its greatest disadvantages are relative insensitivity and difficulty of performance In essence, it measures the adrenal production of corticosterone ("substance B") by the hypophysectomized rat in response to intravenous injection of an unknown sample, and compares this with the response caused by injection of known quantities of reference standard ACTH The corticosterone can be measured in extracts of the adrenal gland, in adrenal vein blood, or in peripheral blood, using the fluorometric technique The second method is the ACTH radioimmunoassay (Yalow et al , 1964, Orth et al , 1968, Berson and Yalow, 1968), which has the great advantage of sensitivity and therefore requires much less plasma for assay While the radioimmunoassay is an extremely useful technique, it has the intrinsic disadvantage that it does not necessarily measure biologically active ACTH, it may also detect inactive ACTH fragments (Besser et al , 1971) Recently, a new assay method has been reported that appears to have many of the advantages of both previous methods In the radioreceptor assay (Lefkowitz et al , 1970), the ACTH receptor site from the cell membrane of an ACTH-responsive adrenocortical tumor cell is used to bind labeled ACTH, rather than using an antibody If the biological activity of an adrenocorticotropin is proportional to its energy of binding to the receptor site, then the competition for binding by unknown samples should give an accurate estimate of the biological ACTH activity of the sample Furthermore, this very sensitive technique should be applicable to any hormone for which a target tissue receptor can be isolated

REGULATION OF MINERALOCORTICOID SYNTHESIS

Renin-Angiotensin System. The synthesis and secretion of adrenal mineralocorticoid are dependent primarily upon the renin-angiotensin system This system is discussed in detail elsewhere in this volume Renin is an enzyme produced by the juxtaglomerular cells which surround the afferent arteriole just before it reaches the glomerular capillary tuft These cells may function as baroreceptors and release renin when renal perfusion falls due to decreased blood pressure or blood volume (Tobian, 1960) Several other factors, including urine sodium concentration at the level of the macula densa in the distal renal tubule and renal sympathetic nervous tone may modify renin release (Vander and Miller, 1964, Gordon et al , 1967, Brubacher and Vander, 1968) Renin converts circulating renin substrate to the decapeptide, angiotensin I A converting enzyme that is produced in the lung removes the two C-terminal amino acids of angiotensin I to form the biologically active octapeptide, angiotensin II Angiotensin II has two major effects *1)* it has an immediate effect on the peripheral arterioles to cause vasoconstriction and increased systemic blood pressure, *2)* it stimulates secretion of aldosterone by the adrenal cortex Aldosterone secretion results, secondarily, in sodium retention, expansion of the intravascular volume, and increased blood pressure Thus, both the acute and chronic effects of angiotensin II tend to correct the original stimulus to renin release, decreased renal perfusion This is another example of a negative feedback regulatory system Angiotensin II is rapidly inactivated by proteolytic enzymes in the plasma and tissues The exact mechanism by which angiotensin II exerts its effect on the adrenal zona glomerulosa cell is unknown It is presumed that it accelerates the conversion of cholesterol to 20α-hydroxycholesterol, the rate-limiting step in aldosterone biosynthesis It is known at least that its site of action precedes the synthesis of DOC (Brown et al , 1971)

Plasma renin activity can be bioassayed in the nephrectomized rat (Boucher et al , 1964) The plasma of the patient is first incubated so as to convert the renin substrate to angiotensin I This is then injected into the rat, where it is converted to angiotensin II The

amount of plasma renin activity is measured by the degree of rise in blood pressure which it causes in the assay animal Plasma renin activity can also be measured by radioimmunoassaying the angiotensin I generated during the incubation of the patient's plasma (Catt et al , 1967, Boyd et al , 1967, Vallotton et al , 1967)

Electrolytes. Increased serum potassium and decreased serum sodium concentrations stimulate aldosterone secretion in the absence of changes in plasma volume in vivo, similar stimulation is observed in vitro Both of these monovalent cations exert their effects early in the aldosterone biosynthetic pathway, prior to the formation of deoxycorticosterone (DOC) (Brown et al , 1971) Aldosterone secretion is sensitive to changes in serum potassium concentration of less than 1 mEq /liter, while a very considerable change in serum sodium is required to effect a similar response

ACTH. The role of ACTH in aldosterone synthesis and secretion is relatively unimportant After chronic hypophysectomy, aldosterone does not rise normally during sodium deprivation It has not been shown that this is due specifically to ACTH deficiency, however In the hypophysectomized rat, ACTH injections do not restore the aldosterone response to hyponatremia (Palmore and Mulrow, 1967) Aldosterone secretion is temporarily increased by very large doses of ACTH, but falls to normal levels despite continued ACTH administration (Liddle et al , 1956)

DISORDERS OF ADRENOCORTICAL FUNCTION

Primary Adrenocortical Insufficiency Addison's disease (Addison, 1855) is usually caused by destruction of the adrenal glands by infection with tuberculosis or histoplasmosis or by idiopathic atrophy associated with circulating antibodies to adrenal tissue (Blizzard and Kyle, 1963) It may also be caused by other granulomatous diseases, infiltration by carcinoma, involvement with amyloid, or by hemorrhage into the adrenals which occurs very rarely in association with severe systemic infections (Waterhouse-Friderichsen syndrome)

The manifestations of primary adrenocortical insufficiency are primarily those of deficient aldosterone secretion, complicated by the loss of cortisol secretion They are quite different, therefore, from the manifestations of secondary adrenocortical insufficiency caused by decreased pituitary ACTH secretion, in which aldosterone production is adequate and only glucocorticoid is deficient Mineralocorticoid insufficiency leads to sodium wasting and potassium retention, causing hyponatremia, hyperkalemia, and a metabolic alkalosis The salt depletion leads to fluid loss and dehydration The effects of glucocorticoid insufficiency are less well-defined They include weakness, irritability, drowsiness and easy fatigability Without the hyperglycemic effects of cortisol to dampen the insulin response to a glucose load, reactive hypoglycemia is frequently observed

The loss of negative feedback inhibition of pituitary ACTH secretion leads to hypersecretion of both ACTH and β-MSH (Fig 7 52)

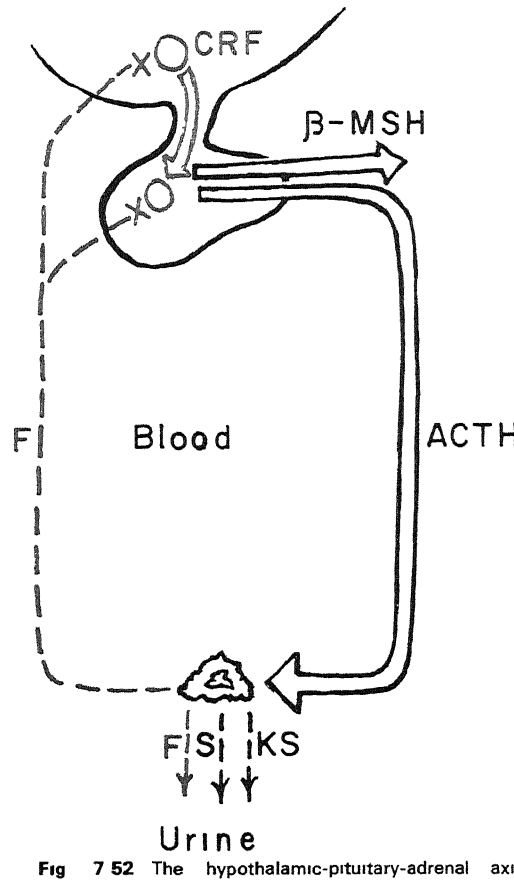


Fig 7 52 The hypothalamic-pituitary-adrenal axis in primary adrenal insufficiency or Addison s disease The atrophic adrenal secretes little or no corticosteroids plasma and urinary steroids are low The loss of cortisol negative feedback inhibition leads to increased secretion of both ACTH and β-MSH by the pituitary

The latter results in "addisonian" hyperpigmentation of the skin, particularly marked in exposed areas and in recent scars, over pressure points such as the elbows, in the buccal mucosa, the areolae, genitals, and intertriginous areas In women, the loss of adrenal androgen results in loss of axillary and pubic hair, men are not affected because of testicular androgen production

The diagnosis can be confirmed by finding low plasma cortisol levels early in the morning, when cortisol is normally high, decreased urinary excretion of 17-OHCS and 17-KS, and failure of the plasma or urinary steroids to increase in response to administration of exogenous ACTH In addition, plasma ACTH and β-MSH are markedly elevated, aldosterone secretion rate is low, and, as a result of decreased plasma volume and renal blood flow, plasma renin activity is increased

Treatment is directed immediately toward restoring plasma volume by infusion of physiological saline solution The addisonian patient requires life-long replacement therapy with both glucocorticoid and mineralocorticoid The adrenal medulla is not necessary for life, no replacement of catecholamines is required for maintenance of good health

Adrenocortical Hyperfunction. Harvey Cushing first described (Cushing, 1932) the "constellation of clinical and metabolic abnormalities resulting from a chronic excess of cortisol" (Liddle, 1967) Apart from the administration of excessive amounts of glucocorticoid medications, there are now three well-defined causes of Cushing's syndrome

1) Adrenal Tumors Benign or malignant adrenal tumors may secrete cortisol autonomously, independent of ACTH stimulation (Fig 7.53)

2) Ectopic Production of ACTH by Nonpituitary Tumors A variety of tumors arising in nonpituitary tissues autonomously secrete ACTH and MSH (Liddle et al, 1969) These hormones appear to be similar to, if not identical with, the authentic pituitary hormones (Fig 7 54)

3) Excess Pituitary Production of ACTH The term, "Cushing's disease," is reserved for this variety of Cushing's syndrome, since Cushing's original concept was that the syndrome he described was caused by an abnormality in pituitary function, "pituitary basophilism" In a few cases, pituitary basophil or

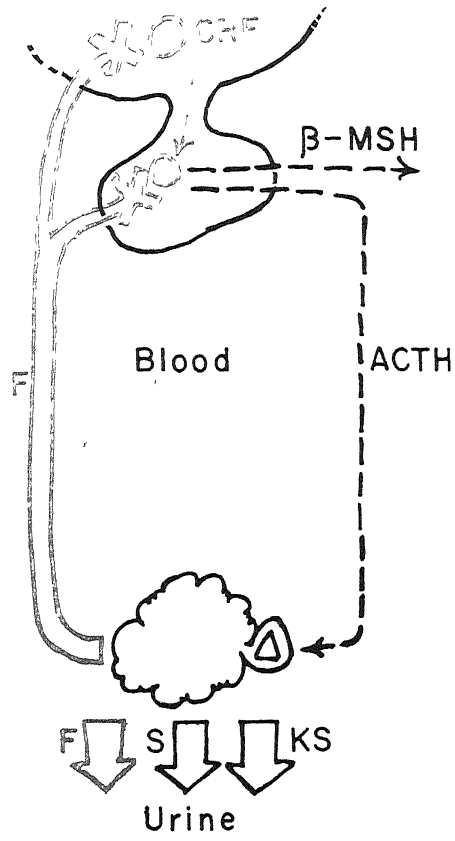


Fig 7 53 Hypothalamic-pituitary-adrenal dynamics in Cushing s syndrome due to adrenocortical tumor The tumor produces increased quantities of cortisol (*F*) deoxycortisol (*S*) and 17-ketosteroids (*KS*) the metabolites of which appear in the urine Plasma cortisol concentrations are increased inhibiting hypothalamic CRF secretion and pituitary ACTH and β-MSH secretion Since ACTH is suppressed the non-tumorous adrenal tissue becomes atrophic

chromophobe tumors may be the source of ACTH In most instances, however, no tumor is found The exact nature of the pituitary aberration is as yet unknown If the defect is at the hypothalamic level, increased secretion of CRF or changes in hypothalamic catecholamine levels might be involved The characteristic feature of Cushing's disease is that ACTH secretion is not autonomous, it is merely "inappropriate" (Liddle, 1967) First, there is relative resistance to glucocorticoid negative feedback, so that it takes a greater than normal plasma cortisol level to suppress ACTH release (Fig 7 55) Second, the normal circadian rhythm in ACTH secretion is lost, whereas, early morning ACTH and cortisol levels may be within the normal range, they do

not fall during the day and are, therefore, clearly elevated above normal values in the evening

The clinical and metabolic abnormalities characteristic of Cushing's syndrome might be predicted from what has already been said about the physiology of cortisol The Cushing's patient is aptly described as "a caricature of the physiologic effects of cortisol" (Rabinowitz, 1968) The patient develops centripetal obesity, wasting of the extremities with increased fat deposition about the trunk, in the supraclavicular areas, the neck and the face There is plethora of the neck and face Wide cutaneous striae are found over the flanks and thighs and, in women, the breasts They represent stretching and thinning of the skin, and

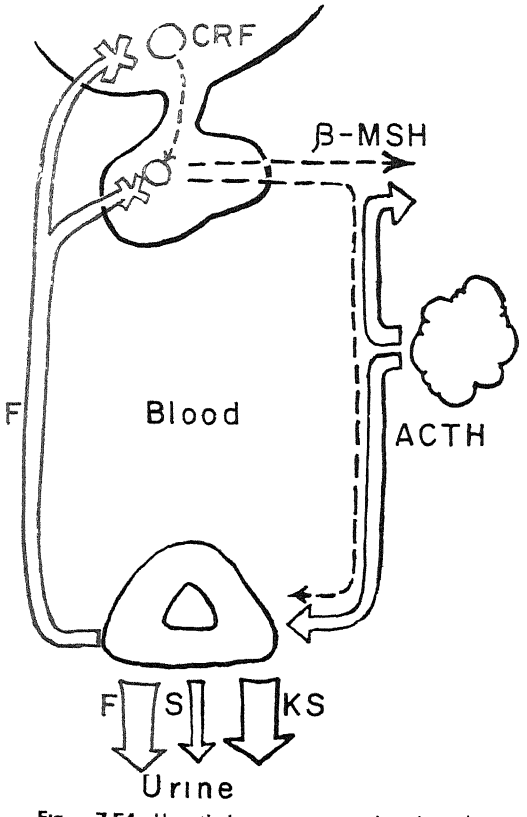


Fig 7 54 Hypothalamic-pituitary-adrenal relationships in the ectopic ACTH syndrome Autonomous hypersecretion of ACTH by a nonpituitary tumor causes bilateral adrenal hyperplasia and increased secretion of corticosteroids Steroid metabolites appear in increased quantities in the urine Plasma cortisol is increased suppressing hypothalamic CRF secretion and pituitary ACTH and β-MSH release Since the tumor itself secretes both ACTH and β-MSH, however, the plasma concentrations of both are increased

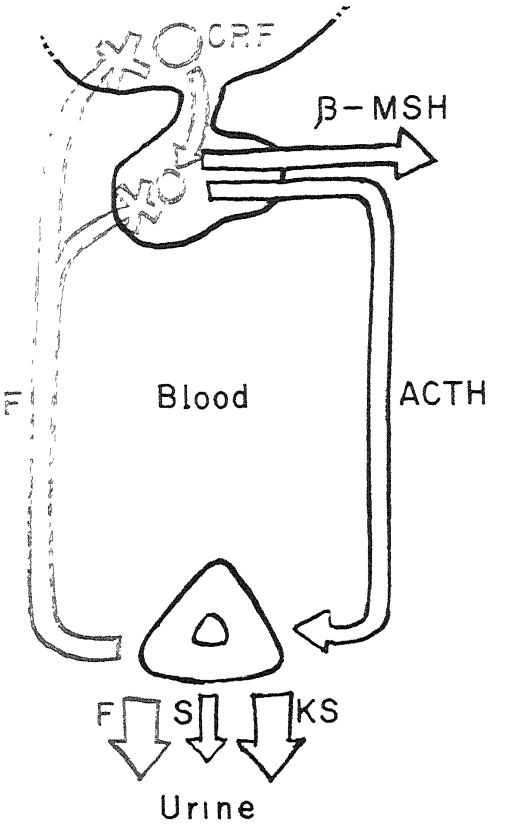


Fig 7 55 Hypothalamic-pituitary-adrenal axis in Cushing s disease The pituitary secretes excessive quantities of ACTH and β-MSH possibly because of increased stimulation by hypothalamic CRF The increased plasma ACTH causes bilateral adrenal hyperplasia and hypersecretion of corticosteroids Urinary corticosteroid levels are increased The hypothalamic-pituitary axis is relatively resistant to suppression since elevated plasma cortisol concentrations fail to suppress pituitary ACTH and β-MSH secretion into the normal range

appear purple because they reveal the underlying vascular dermia Wounds heal poorly The patient's muscles become atrophic, and he cannot rise from a squatting position without assistance Carbohydrate intolerance is the rule, and frank diabetes mellitus may be seen in patients with a familial predisposition to this disorder The patient may have mental aberrations ranging from euphoria to frank schizophrenia In children, a dramatic arrest of linear growth is observed In adults, osteoporosis may be severe, resulting in compression fractures of the vertebrae The sodium-retaining effects of high cortisol levels result in hypertension, which may be aggravated by the potentiation of the vasoconstrictor effects of

endogenous catecholamines The Cushing's patient is abnormally susceptible to infections Menstrual aberrations and downy hirsutism are frequent findings in women, as is acne over the face and shoulders in both sexes Hyperpigmentation may be seen in Cushing's disease or the ectopic ACTH syndrome because of the effects of β-MSH

The diagnosis of Cushing's syndrome can be confirmed and its cause determined by applying physiology to clinical medicine (Fig 7 56) In the patient with suspected Cushing's syndrome, the physician tests his clinical impression by determining if hypercortisolism is actually present If plasma concentrations and urinary excretion of 17-OHCS are elevated, if there is no diurnal rhythm in plasma cortisol concentration, and if ACTH (and, secondarily, cortisol) secretion cannot be suppressed by low doses of the potent glucocorticoid, dexamethasone, the patient has Cushing's syndrome The cause of the hypercortisolism can then be determined by two further procedures We have already seen that ACTH secretion in Cushing's disease is only relatively resistant to feedback inhibition Thus, pituitary ACTH secretion will be suppressed with high doses of dexamethasone If it is not, the patient has either autonomous production of ACTH by a nonpituitary tumor or autonomous production of cortisol by a benign or malignant adrenal tumor These two can be differentiated by measuring plasma ACTH In the ectopic ACTH syndrome, plasma ACTH is obviously elevated In the patient with Cushing's syndrome due to adrenal tumor, on the other hand, autonomous hypersecretion of cortisol suppresses ACTH secretion by the normal pituitary, and plasma ACTH is low

Treatment of Cushing's syndrome depends on the cause If there is a tumor, the only satisfactory treatment is surgical removal In Cushing's disease, one can either direct one's attack at the pituitary, to reduce ACTH secretion, or at the adrenals, to remove the source of the excess cortisol Pituitary irradiation is successful in reducing ACTH secretion to normal in many patients Surgical hypophysectomy has also been used Bilateral surgical adrenalectomy is successful in curing Cushing's disease, but the patient then requires steroid replacement therapy for the rest of his life Adrenolytic drugs and drugs which interfere with corticosteroid biosynthesis are also sometimes used (Temple and Liddle, 1970)

Congenital Adrenal Hyperplasia. The first detailed description of this syndrome (Baulieu et al, 1967) was published by DeCrecchio (1865) It is caused by inborn errors of metabolism involving deficiency of specific enzymes which participate in adrenal steroid biosynthesis The common feature of the congenital adrenal hyperplasias is that cortisol synthesis is impaired (Fig 7 57), resulting in increased

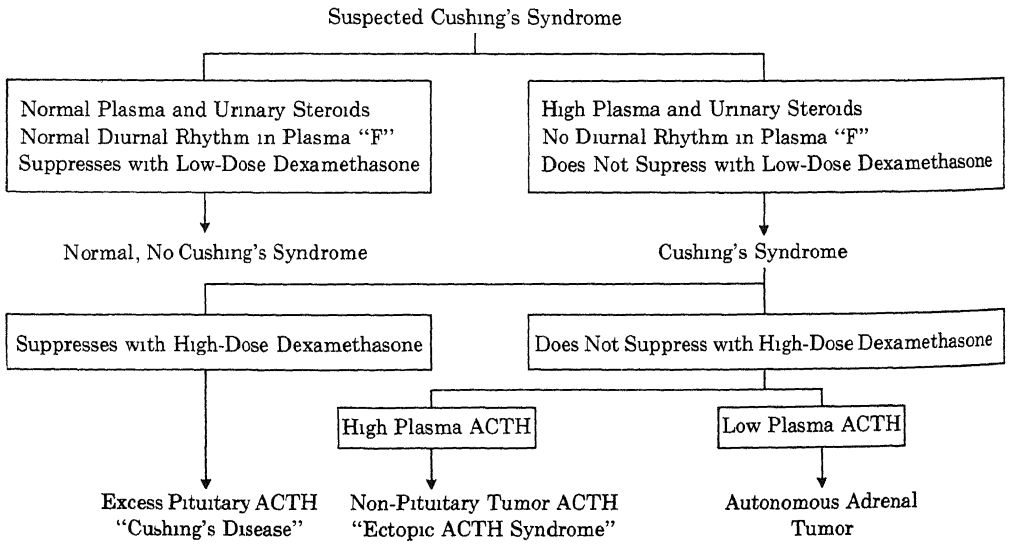


Fig 7 56 A flow diagram for establishing the diagnosis of Cushing s syndrome and determining its etiology by means of physiologic manipulations See text for discussion

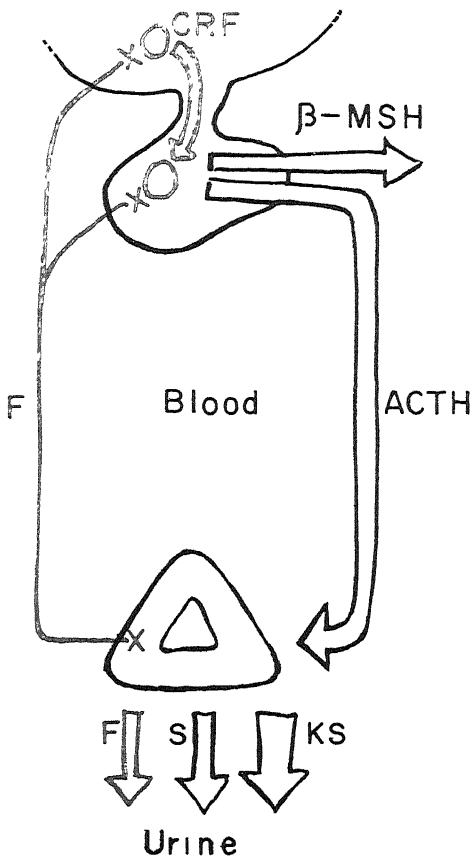


Fig 7 57 Hypothalamic-pituitary-adrenal dynamics in the congenital adrenal hyperplasias The adrenal cannot synthesize normal amounts of cortisol because of an inborn enzymatic defect Low plasma cortisol concentrations result in increased secretion of ACTH and β-MSH by the pituitary Under the influence of chronic ACTH stimulation the adrenals become increasingly hyperplastic and cortisol precursors are produced in increasing quantities until normal or near-normal plasma cortisol levels are achieved The cortisol is produced at the expense of increased secretion of androgenic 17-ketosteroid precursors the metabolites of which appear in the urine

ACTH secretion by the pituitary Increased ACTH stimulation results in bilateral adrenal hyperplasia, increasing the total adrenal mass and providing sufficient precursor steroids that cortisol production reaches normal or near-normal levels Normal cortisol production is achieved at the expense of excessive production of adrenal androgens, however, which cause the characteristic virilization seen in this disorder There are four major types of congenital adrenal hyperplasia *1)* simple virilizing, *2)* salt-losing, *3)* hypertensive, and *4)* 3β-ol dehydrogenase defect, in order of decreasing incidence There are, in addition, other very rare forms that will not be discussed here

Simple Virilizing Type A partial deficiency of the 21-hydroxylase enzyme is responsible for this type of congenital adrenal hyperplasia, more complete deficiency causes the salt-losing variety

The defect in steroid biosynthesis is shown in a simplified schematic diagram in Figure 7 58 Synthesis of both cortisol and aldosterone are affected Subnormal cortisol production results in increased plasma ACTH levels, which cause accelerated production of 20α-hydroxycholesterol and, since this is the rate-limiting step in steroidogenesis, increased production of the other cortisol precursors Cortisol levels are normal or subnormal in both plasma and urine, depending upon the degree of enzyme deficiency Urinary excretion of 17-ketosteroids (DHA and its sulfate, Δ^4-androstenedione, androsterone, and etiocholanolone) is increased Metabolites of cortisol precursors also appear in increased quantities in the urine Pregnanediol is a metabolite of pregnenolone and progesterone, and pregnanetriol is derived from 17-hydroxypregnenolone and 17-hydroxyprogesterone Pregnanediol and pregnanetriol are easily measured in urine (Bongiovanni and Clayton, 1954) Testosterone is also increased in the urine It is not measured in the 17-KS, but must be quantitated by other methods (Camacho and Migeon, 1966) Since the adrenals are already maximally stimulated by endogenous ACTH, plasma and urinary steroids are usually not increased by administration of ACTH

The clinical picture depends upon the severity of the enzyme defect, which dictates the degree of excess androgen production, and the time at which excessive androgen secretion begins during fetal life In the female, pseudohermaphroditism results, with varying degrees of genitourinary anomaly Increased pigmentation, particularly of the areolae and the genitals, may be observed, and is the result of increased β-MSH secretion If the patient is not treated, progressive virilization will occur, with increasing clitoral hypertrophy and development of a male pattern of pubic hair, usually be the age of 3 years Early growth acceleration is accompanied by early closure of the epiphyseal growth centers Although these patients are tall as young children, they do not reach normal adult height if untreated The

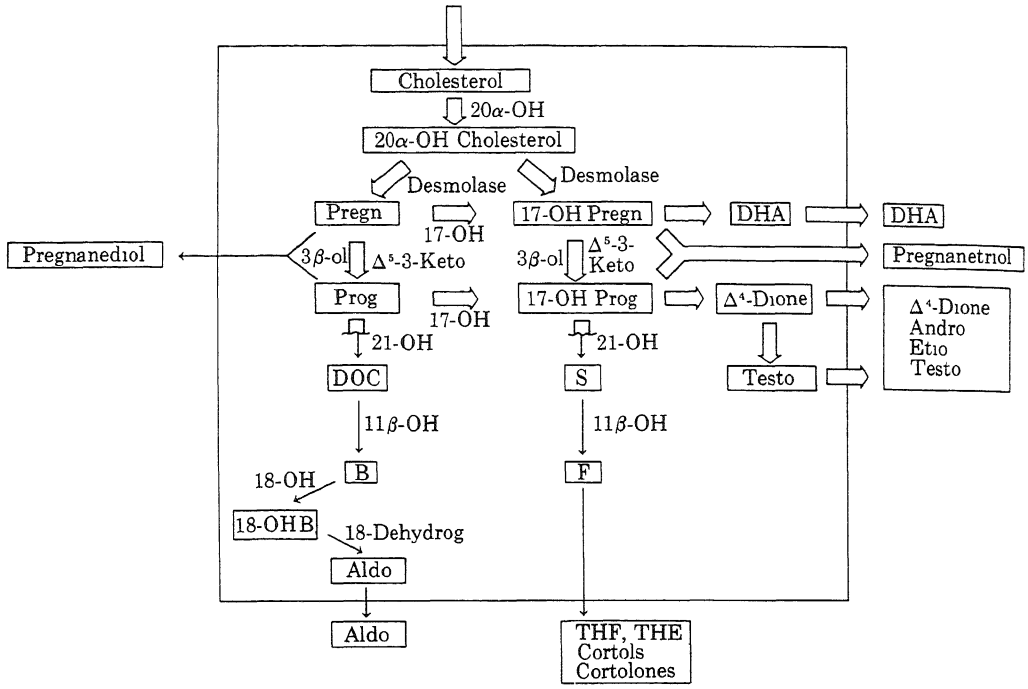


Fig 7 58 A diagram of the biosynthetic defect in simple virilizing congenital adrenal hyperplasia and salt losing congenital adrenal hyperplasia both of which are the result of a deficiency in 21-hydroxylase Refer to Figure 7 44 and to the text for discussion

child may be raised as a boy and will develop a masculine body habitus, deep voice, body hair, and beard The breasts fail to develop, the uterus remains infantile, and the patient fails to menstruate

The syndrome in the male is much more difficult to recognize Except for increased pigmentation, no abnormalities are obvious at birth Precocious puberty and rapid growth—macrogenitosomia praecos—usually are noted within the first few years While the child may appear to be a "young Hercules," his epiphyses, too, will close at age 8 or 10 if he is not treated, and normal adult stature will not be achieved In contrast to enlargement of the penis and prostate, due to stimulation by adrenal androgens, the testes usually remain small and immature, since they require stimulation by pituitary gonadotropic hormones

The diagnosis depends upon the early recognition of the syndrome and is verified by finding normal or subnormal 17 OHCS levels in plasma and urine and markedly elevated 17-KS and pregnanetriol The markedly elevated urinary levels of pregnanetriol and other metabolites of 21-deoxy steroids causes the 17-KGS to be increased Thus, they do not reflect the rate of cortisol secretion in this condition The 17-KS (and the 17-KGS) rapidly suppress to normal with replacement doses of glucocorticoids, since pituitary ACTH is normally responsive to feedback inhibition

Treatment consists only of replacement corticosteroid therapy The diagnosis must be made at an early age in order to prevent growth arrest and, in the female, in order to prevent progressive virilization and male psychosexual fixation

Salt-Losing Type. This represents a more severe form of the simple virilizing type of congenital adrenal hyperplasia The signs and symptoms are the same, with the additional very serious complication of the syndrome of primary adrenal insufficiency This usually manifests itself in dehydration and shock within the first 2 weeks of life, but may occur later in childhood, precipitated by some stress such as infection Death may occur because of hypovolemic shock or cardiac arrest caused by extreme hyperkalemia

There are at least three factors involved in the salt wasting (Baulieu et al , 1967) First, aldosterone secretion does not rise normally with salt deprivation, there is loss of aldoster-

one secretory "reserve" (Blizzard et al., 1959) Second, cortisol normally exerts a permissive effect on the renal tubular actions of aldosterone (Eberlein and Bongiovanni, 1958), and cortisol secretion is impaired Third, progesterone and 17-hydroxyprogesterone, which are secreted in very large quantities in this syndrome, antagonize the actions of aldosterone, perhaps by competing for intracellular receptor-binding sites (Crigler et al 1952, Jailer, 1953) These patients usually require mineralocorticoid replacement therapy

Hypertensive Type. Much less common than the 21-hydroxylase deficiency types, this variety is caused by a deficiency in 11β-hydroxylase (Fig 7 59) The pathogenesis and clinical syndrome are essentially the same, except for the associated moderate to severe high blood pressure The urinary 17-OHCS are actually elevated in this syndrome because of increased excretion of tetrahydrodeoxycortisol (THS), the last steroid precursor before the enzymatic defect Plasma and urinary cortisol and cortisol metabolites are normal or low, depending upon the completeness of the enzyme deficiency The hypertension is caused by hypersecretion of deoxycorticosterone (DOC), a moderately potent mineralocorticoid (Table 7 5) The hypertension is corrected and the virilization arrested by simple glucocorticoid replacement therapy

Another variety of hypertensive congenital adrenal hyperplasia has recently been described in which the activity of the 17-hydroxylase enzyme is deficient (Biglieri 1966) Since this deficiency is present not only in the adrenal but in the ovary, neither adrenal androgens nor ovarian estrogens are synthesized Increased synthesis of corticosterone and deoxycorticosterone causes hypertension and hypokalemic alkalosis, aldosterone secretion is actually suppressed by the salt-retaining effects of deoxycorticosterone The patient is phenotypically female, but fails to menstruate or sexually mature because of lack of estrogen The diagnosis is confirmed by finding absent 17-OHCS, 17-KGS, and 17-KS in the urine in the presence of excessive quantities of tetrahydrocorticosterone and tetrahydrodeoxycorticosterone Treatment consists of replacement both with glucocorticoid and with estrogen and progesterone

3β-ol Deficiency Type This rare form of congenital adrenal hyperplasia is usually fatal early in infancy because of severe sodium loss and resulting dehydration and hypovolemic

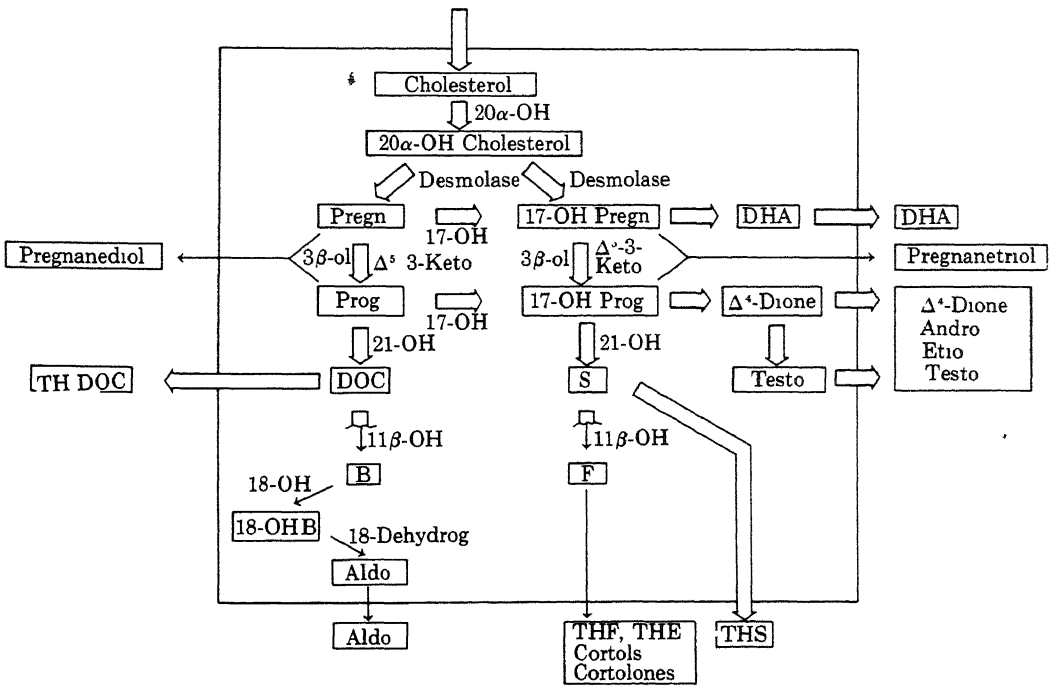


Fig 7 59 The biosynthetic defect in the hypertensive form of congenital adrenal hyperplasia caused by 11β-hydroxylase deficiency Refer to Figure 7 44 and to the text for discussion

shock Essentially no physiologically useful adrenal corticosteroids are synthesized Virilization of females is less marked than in the other congenital adrenal hyperplasias because no androgen more potent than DHA is secreted In addition, males very often have incomplete fetal genital development, with severe hypospadias Since fetal testicular androgen is required for normal male genital development, it appears that the testes share the 3β-ol dehydrogenase deficiency (Bongiovanni, 1962) These patients, if diagnosed in time, require complete steroid replacement Even then, few survive more than a month or two It is not known why these infants fail to respond to apparent adequate replacement of their missing steroid hormones

Primary Hyperaldosteronism. Autonomous hypersecretion of aldosterone by an adrenocortical tumor arising from the zona glomerulosa causes a syndrome due to pure mineralocorticoid excess (Conn, 1955) The major features of the syndrome are hypertension and hypokalemic alkalosis The hypertension is secondary to sodium retention and, possibly, to partial obliteration of the lumina of peripheral arterioles by swollen arteriolar cells These patients do not develop edema, despite the fluid retention, unless congestive heart failure occurs Hypokalemia is often mild, but may be worsened by diuretic therapy Severe hypokalemia may result in weakness, paresthesias, decreased deep tendon reflexes, cardiac arrhythmias, and even paralysis The hypokalemia interferes both with insulin secretion and with the peripheral action of insulin Carbohydrate intolerance is seen in half the patients with this syndrome The low serum potassium also interferes with renal concentrating ability, polyuria and polydipsia are the result The urine is dilute and neutral

The diagnosis is confirmed by finding of *1)* an elevated aldosterone secretion rate on a liberal salt diet (100 mEq sodium per day) which tends to suppress aldosterone secretion by increasing renal perfusion, and *2)* a suppressed plasma renin activity, even after sodium depletion (10 mEq sodium diet and sodium-wasting diuretic therapy) and during upright posture Furthermore, if balance studies are performed, it will be found that the patient is in negative potassium balance and positive sodium balance Treatment is removal of the aldosterone-secreting adrenal tumor Aldosterone antagonists, such as the spironolactones are often useful in correcting the electrolyte abnormalities and hypertension prior to surgery

Secondary Hyperaldosteronism. Hyperaldosteronism secondary to increased stimulation by the renin-angiotensin system occurs when perfusion of the kidneys, or a portion of one of them, is decreased This is a normal, homeostatic defense mechanism when effective arterial plasma volume is decreased This occurs in dehydration, hemorrhage, or shock It may also occur when intravascular fluid is lost to another compartment in the body the interstitial spaces filled with edema fluid in congestive heart failure, or the intraperitoneal space filled with ascitic fluid in hepatic cirrhosis In these situations, the renin-angiotensin-aldosterone system is functioning normally to maintain effective arterial plasma volume There is a condition, however, in which secondary aldosteronism is pathological In unilateral renal artery stenosis, the main renal artery or one of its segmental branches may become partially occluded so that the renal cortex supplied by it becomes ischemic The juxtaglomerular cells in this portion of the kidney then behave as though arterial volume were decreased, they release renin The renin-angiotensin system stimulates aldosterone secretion, and a syndrome that mimics primary aldosteronism develops It can be differentiated from primary hyperaldosteronism, however, by demonstrating both an elevated aldosterone secretion rate and an elevated plasma renin activity Assay of plasma renin activity in blood obtained by catheterizing each renal vein will reveal greatly elevated levels on the side of the stenotic artery (Michelakis et al , 1967)

Treatment is directed at relieving the stenosis by vascular surgery or, if that is not possible, removal of the ischemic kidney

The Adrenal Medulla

DEVELOPMENT AND ANATOMY

Early in fetal life, sympathetic cell precursors migrate from the neural tube which eventually differentiate into either sympathetic ganglion cells or chromaffin cells The latter cells remain closely associated with the sympathetic nervous system and form the paraganglia, the preaortic plexuses of Kuhn and

Zuckerkandl, the carotid and glomus bodies, and the adrenal medulla Similar chromaffin cells are associated with the sympathetic and parasympathetic nerves in many organs and viscera of the body While most of these cells atrophy during infancy, some survive and can be the origin of "ectopic" functional hormone-secreting tumors in later life

The medulla is composed at birth of primitive sympathetic cells which fully differentiate into chromaffin cells (pheochromocytes) during the first 3 years of life Most of the blood supply to the adrenal medulla arrives via a portal system draining the cortex Thus, the afferent blood has very high concentrations of adrenal corticosteroids In contrast to the cortex, the adrenal medulla is not essential to life, although it normally functions to release catecholamines in the "fight or flight" response to emotion or stress

BIOSYNTHESIS OF CATECHOLAMINES

The synthesis of epinephrine, the important catecholamine secreted by the adrenal medulla, is outlined in Figure 7 60 Tyrosine is derived primarily from dietary sources (Levine and Conn, 1967) The first step in the biosynthetic pathway, conversion of tyrosine to dopa, is the rate-limiting reaction (Levitt et al , 1965, Kaufman and Friedman, 1965) The reaction is catalyzed by tyrosine hydroxylase, an enzyme found only in neurectodermal tissues Tyrosine hydroxylase is inhibited by a variety of catecholamines, which thereby function as an intracellular negative feedback regulatory control on their own synthesis (Undenfriend, 1968) Dopa is converted to dopamine by L-aromatic amino acid decarboxylase, and thence to norepinephrine by dopamine β-oxidase Phenylethanolamine-N-methyl trans-

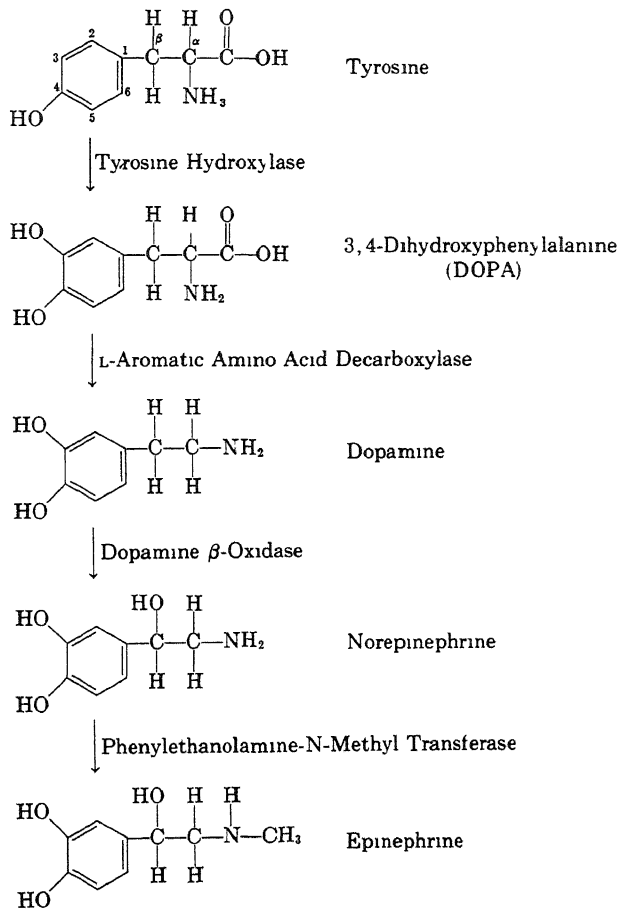


Fig 7 60 The biosynthetic pathway for conversion of tyrosine to norepinephrine and epinephrine

ferase (PNMT), the enzyme that catalyzes the conversion of norepinephrine to epinephrine, is found in significant concentrations only in the adrenal medulla This is of interest because phenylethanolamine-N-methyl transferase is induced by high concentrations of cortisol, which are found only in the portal blood draining the adrenal cortex and supplying the medulla (Wurtman and Axelrod, 1966) Thus, while other catecholamines are synthesized widely in the central nervous system and in sympathetic nerve endings, epinephrine is synthesized almost exclusively by the adrenal medulla

STORAGE AND RELEASE OF CATECHOLAMINES

Unlike the adrenocortical steroids, which are released almost as soon as they are synthesized, catecholamines are stored in the adrenal medullary cells They are held in membrane-enclosed secretory vesicles similar to those seen in the synaptic nerve endings of the sympathetic nerves They are complexed with adenosine triphosphate (1 mole of ATP for 4 moles of catecholamine) and a specific protein (MW 40,000) The vesicles also contain a Ca^{++} and Mg^{++} dependent ATPase (Kopin, 1967) The binding of the amine to the complex is moderately structure-specific and requires both the catechol configuation (Fig 7 61) and the β-hydroxyl group for maximum binding These subcellular vesicles take up epinephrine and norepinephrine by a transport process requiring energy The free catecholamine which is not taken up immediately by the vesicles is metabolized by monamine oxidase (MAO) located within the mitochondria The hormone contained within the secretory vesicle can be released in response to an appropriate stimulus which, in the case of the adrenal pheochromocyte, is acetylcholine secreted by the preganglionic nerve fibers that terminate in the medulla Other agents, such as histamine and glucagon, can stimulate epinephrine release, their physiologic importance is not clear Calcium ions enter the cell in response to acetylcholine stimulation and mediate the excitation-secretion coupling phenomenon by which a unit stimulus causes the release of a "quantum" of catecholamine The quantal release of epinephrine is due to the fact that the contents of a single secretory vesicle are extruded into the extracellular space The catecholamine, vesicle-binding protein and ATP products are released in the same ratio in which they are found in the secretory vesicle (Douglas, 1966a, 1966b, Kirschner et al , 1966) In the synapses of the sympathetic nerve cells, most of the norepinephrine released is taken up immediately by the cell and incorporated into secretory vesicles There is little evidence to suggest that this is true of the adrenal medullary pheochromocyte Presumably, most of the epinephrine is carried away either to act upon distant target tissues or to be metabolized by catechol-O-methyl transferase (Fig 7 62)

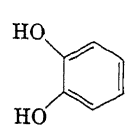


Fig 7 61 Catechol (pyrocatechin)

METABOLISM OF CATECHOLAMINES

Norepinephrine and epinephrine are metabolized by two enzymes, primarily Monamine oxidase (MAO), which is found in the mitochondria of chromaffin cells (DeLores Arnaiz and DeRobertis, 1962), deaminates the α-carbon of norepinephrine, epinephrine, and other catecholamines, such as serotonin and dopamine The free epinephrine not contained within the secretory vesicles is metabolized in situ by this enzyme Catechol-O-methyl transferase (COMT) catalyzes the transfer of the methyl group from S-adenosyl methionine to the catechol 3-hydroxyl radical and requires Mg^{++} as a cofactor (Axelrod, 1957, Wurtman, 1965) Catechol-O-methyl transferase is present in many tissues and metabolizes the circulating free catecholamines The major metabolites of epinephrine and norepinephrine are shown in Figure 7 63

ASSAY OF CATECHOLAMINES

The technique for assaying catecholamines requires that they first be extracted from the specimen either by adsorption onto aluminum hydroxide at alkaline pH, followed by elution with weak acid, or by ion exchange chromatography on acid resins They are then converted to fluorescent derivatives which can be quantitated The most commonly used method involves oxidation in weak acid to form the qui-

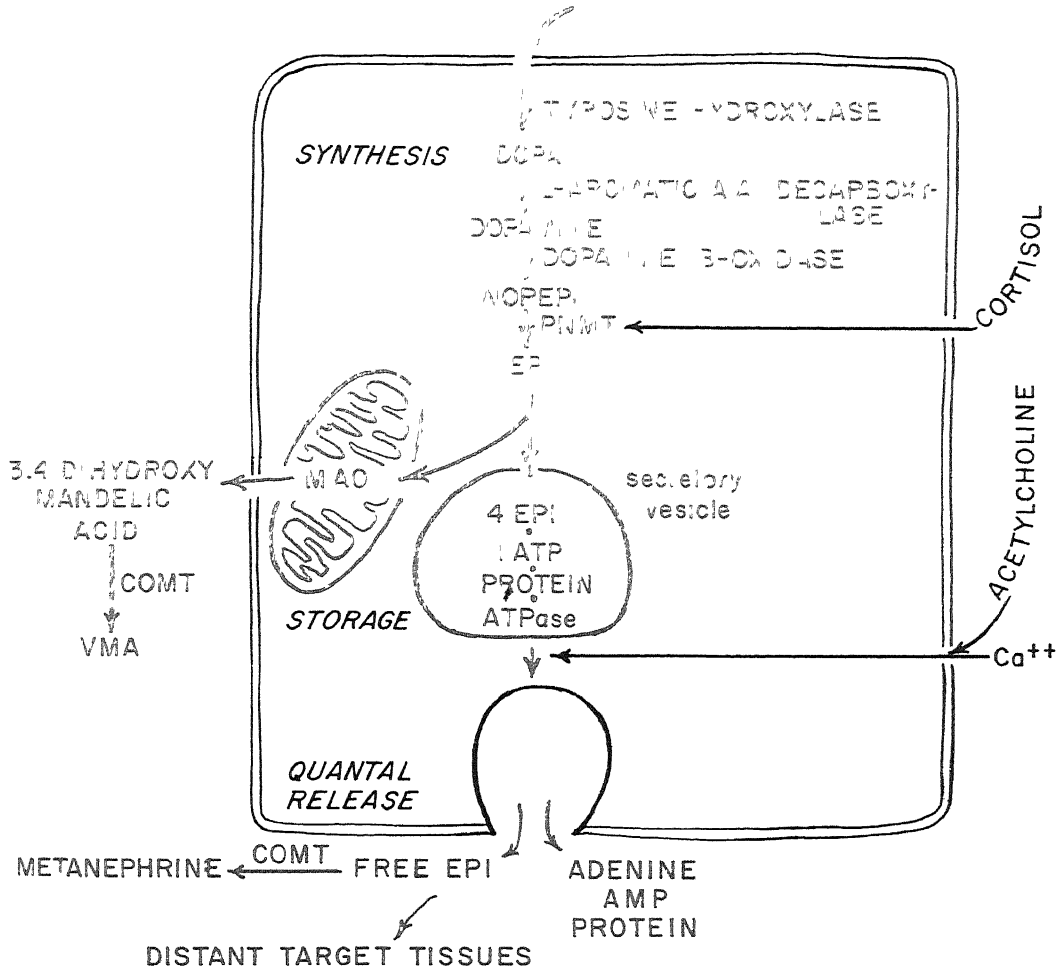


Fig 7 62 A diagrammatic representation of epinephrine synthesis storage release and metabolism by the adrenal medullary pheochromocyte See text for discussion

none derivatives (adrenochrome, noradrenochrome) with subsequent conversion, by exposure to alkaline solution, to their fluorescent trihydroxyindole derivatives (Fig 7 64) (Anton and Sayre, 1962, Haggendal, 1963) The method is quite specific, dopa and dopamine interfere somewhat, but they are usually present in only small quantities. The antihypertensive agent, α-methyl dopa, on the other hand, interferes grossly in this assay

Free or conjugated catecholamines, vanillyl mandelic acid (VMA), or total metanephrines (normetanephrine plus metanephrine) can be assayed in the urine (Page and Copeland, 1968) Free epinephrine is excreted in the amount of 20 μg or less per day The excretion of free norepinephrine, the great majority of which is released by the sympathetic nerve endings, is 80 μg or less per day The upper limits of normal for total metanephrines and VMA are 1 3 mg and 6 5 mg /day, respectively

PHYSIOLOGICAL EFFECTS OF CATECHOLAMINES

The catecholamines exert their effects on target tissues by binding to a receptor site on the cell membrane, influencing the activity of the associated enzyme, adenyl cyclase (Sutherland and Rall, 1960) Adenyl cyclase regulates the synthesis of cyclic 3′,5′-AMP Cyclic AMP is the intracellular mediator of catecholamine action The receptors have been divided into two categories by pharmacologists on the basis

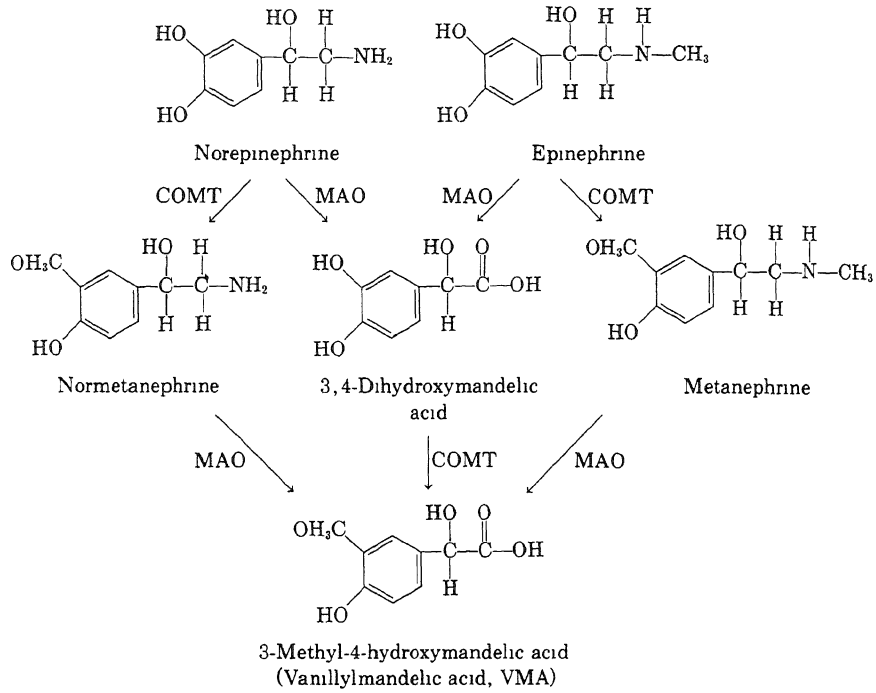


Fig 7 63 The major pathways of metabolism of norepinephrine and epinephrine Two enzymes are involved Catechol-O-methyl transferase (COMT) and monamine oxidase (MAO)

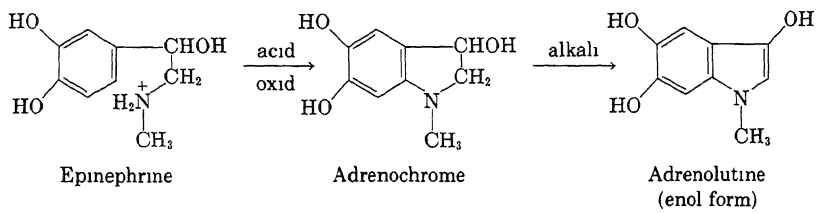


Fig 7 64 Conversion of epinephrine to adrenolutine a fluorescent trihydroxyindole in the assay of catecholamines

of the compounds that bind to the receptor sites (Ahlquist, 1948). Although there is no compelling evidence for the existence of two structurally distinct receptors, an α (vasoconstrictor) and β (vasodilator) receptors have been identified Since synthetic agents have been developed that stimulate or block only one of these types of activities, however, it has remained a useful working concept (Sutherland and Rall, 1960) There is some evidence that β-adrenergic agents increase intracellular cyclic AMP levels and that α-adrenergic agents decrease it (Abe et al, 1969b), but the exact mechanism of α-adrenergic agents is still in doubt Both norepinephrine and epinephrine have both α- and β-adrenergic activities

A list of some of the metabolic effects of catecholamines is given in Table 7 6 Catecholamines stimulate glycogenolysis and therefore have an anti-insulin effect The catecholamines stimulate the mobilization of free fatty acids from adipose tissue, skeletal muscle and myocardium in the presence of minimal concentrations of glucocorticoids, which exert a permissive effect These actions, which involve activation of adenyl cyclase, might be expected to be β-receptor activities This appears to be the case Epinephrine is more po-

Table 7 6
Some effects of catecholamines

Effector Organ	Receptor Type	Response
Eye		
Radial muscle, iris	α	Contraction (mydriasis)
Ciliary muscle	β	Relaxation for far vision
Heart		
Sinoatrial node	β	Increase in heart rate
Atrioventricular node	β	Increase in conduction velocity and shortening of functional refractory period
Atria	β	Increase in contractility
Ventricles	β	Increase in contractility and irritability
Blood vessels	α	Constriction
	β	Dilatation (predominates in skeletal muscle)
Bronchial muscle	β	Relaxation (bronchodilatation)
Gastrointestinal tract		
Motility		
Stomach	β	Decrease
Intestine	α, β	Decrease
Sphincters	α	Contraction
Urinary bladder		
Detrusor	β	Relaxation
Trigone and sphincter	α	Contraction
Skin		
Pilomotor muscles	α	Piloerection
Sweat glands	α	Selective stimulation (adrenergic sweating)

From Levine (1969)

tent than norepinephrine in producing these effects

DISORDERS OF ADRENOMEDULLARY FUNCTION PHEOCHROMOCYTOMA

Pheochromocytomas are functioning tumors, usually benign, that arise from pheochromocytes They occur most commonly in the adrenal medulla, but may arise wherever these chromaffin cells are found in the body They occasionally are familial (Page and Copeland, 1968) and may be associated with benign tumors of other endocrine glands, malignant tumors of the thyroid, and certain other conditions (von Recklinghausen's neurofibromatosis, von Hippel-Lindau's hemangioblastomatosis)

The clinical manifestations of pheochromocytoma are due to catecholamine hypersecretion and depend upon whether the tumor tends to release the catecholamines constantly or paroxysmally, and whether it secretes mostly epinephrine or norepinephrine Hypertension, usually constant but sometimes paroxysmal and associated with excessive perspiration, flushing, palpitations, and tremor, is the hallmark of the disease The patient may also have headaches, pallor, nausea, nervousness, and weakness

The diagnosis is confirmed by documenting increased excretion of catecholamines, VMA or total metanephrines in the urine If epinephrine is elevated without a parallel increase in norepinephrine, the tumor is virtually always found in the adrenal medulla of the organ of Zuckerkandl Treatment is removal of the tumor

7 / HORMONAL CONTROL OF THE PANCREATIC ISLETS

Insulin and Glucagon

Insulin and glucagon are peptide hormones which are secreted by pancreatic islet cells and exert important regulatory effects on carbohydrate, fat, and protein metabolism The discovery of insulin is closely linked with attempts to understand the nature of diabetes mellitus, a disorder which results from a relative or absolute deficiency of this hormone Studies of von Mehring and Minkowski in the late 1800's, clearly established the pancreatic origin of diabetes mellitus by showing that the disorder appeared in dogs following pancreatectomy In 1922, Banting and Best reported the isolation of an active hormonal principal, insulin, from the pancreas and were able to reverse the metabolic changes of diabetes In the early 1950's, Sanger and co-workers (Sanger, 1959) established the complete amino acid sequence of insulin This pioneering accomplishment ushered in a new era in protein chemistry and led to the equally exciting achievement of the total chemical synthesis of a biologically active protein Vasopressin and oxytocin were the first peptide hormones to be synthesized, and the list now includes insulin (Katsoyannis et al , 1964) and glucagon (Wunsch, 1967)

CHEMICAL NATURE OF INSULIN

Insulin has a molecular weight of about 6000 and consists of two polypeptide chains, A and B, which are joined at two places by disulfide bridges (—S—S—) of cystine (Fig 7 65) A third disulfide bridge within the A-chain forms an intrachain disulfide ring Similar cyclic disulfide structures are found in vasopressin and oxytocin but, as far as is known, these two posterior lobe hormones do not share any important physiological property of insulin

Insulin activity is thought to be related closely to its natural conformation or three-dimensional shape, which in turn is determined largely by its primary or amino acid structure The three-dimensional shape of insulin is stabilized by hydrogen bonding and loss of biological activity occurs when these bonds are broken (denaturation) as by excessive heat or exposure to alkali, even though there has been no alteration in amino acid structure The disulfide bridges are also important biologically and their disruption by reducing agents will result in nearly complete loss of insulin potency On the other hand, many chemical modifications of the amino acid structure of insulin have been made without significant alteration in biological activity Some of these alterations have proved to be useful in labeling insulin with radioactive markers One of the most widely used modifications at the present time is ^{125}I- or ^{131}I-labeled iodoinsulin By its use, it has been possible to measure very small concentrations of insulin in metabolic experiments either directly or by the use of radioimmunoassay procedures In addition to these man-made modifications, many natural variations in amino acid composition of native insulin have been found In a comparison of amino acid sequences among 16 or so species, variations have been found in about 60% of the amino acid residues (Smith, 1966) One unchanging feature, however, is the relative position of the disulfide bridges, and in all species of insulin thus far identified, identical numbers of amino acids are contained within the cyclic peptide in the A-chain and between the points of attachment of the disulfide bridges joining the A- and B-chains The only differences noted among the five species of mammalian insulin —beef, pork, sheep, horse, and whale—reside in the sequence of three amino acids within the A-chain ring (Smith, 1966) With the exception of guinea pig insulin which exhibits reduced biological potency in other mammals, none of these natural variations in amino acid structure is associated with alteration in biological activity. These variations do confer differences in immunologic activity, however, and it is sometimes necessary in the management

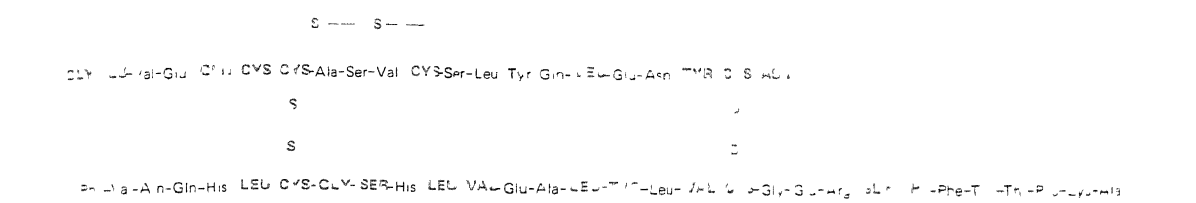


Fig 7 65 Structure of bovine insulin redrawn to illustrate the invariant amino acids (brown type) among vertebrates (Ryle et al 1955 Smith 1966)

of patients with diabetes to change the source of mammalian insulin when insulin resistance from antibodies complicates treatment

INSULIN BIOSYNTHESIS

Pancreatic Islets (Lacy, 1970) Insulin biosynthesis is the exclusive property of the β-cell of pancreatic islet tissue (Fig 7 66) In some lower vertebrates, notably certain teleosts, islet tissue is formed as a separate organ from the exocrine pancreas, but in higher forms, including all mammals, the islet cells are scattered throughout the pancreas Human pancreatic islets are characterized by the presence of four distinct cell types, each with its own unique type of secretory granules Analysis of the insulin granules indicate that the insulin molecules are packed as a structural unit giving rise to a distinct crystalline periodicity Zinc appears to be bound with this complex but its importance is still not understood The secretory mechanism of β-cells is a matter of debate While many investigators feel that the granule is released by reverse pinocytosis, a process referred to as emiocytosis, other investigators have found no evidence for fusion of the limiting membrane of secretory granules with the plasma membrane of the cell The latter would imply an intracytoplasmic dissolution of the granule and subsequent transport of insulin across the plasma membrane to the interstitial space α-cells are characterized by the presence of extremely electron opaque granules and are thoroughly documented as a site of glucagon secretion in the islet The secretory mechanism of these cells is not known Δ-cells are present in the pancreatic islets of all species studied to date and are characterized by the presence of a secretory granule that has moderate electron opacity and is extremely uniform in its ultrastructural appearance The hormones secreted by Δ-cells are not clearly established although there is some data from immunofluorescence studies suggesting that gastrin is secreted by these cells The ubiquitous fourth cell type which has been described in monkey, human, and rabbit pancreatic islets as well as the islets of some reptiles, appears to resemble some of the enterochromaffin cells of the intestinal tract

Proinsulin The amino acids in insulin are presumed to be assembled in peptide linkage on ribosomes in the same way that other peptides are formed and the reader is referred to Chapter 13 for a description of this process While our knowledge of protein synthesis is able to explain the formation of single chain peptides, the question as to how the separate A- and B-chains are combined was a mystery until recently It is known from laboratory experiments that the recombination of separated A- and B-chains is often extremely inefficient, since chains often join together in a haphazard fashion producing biologically inactive material The problem was solved by the discovery of proinsulin (Steiner et al, 1969) As shown in Figure 7 67, proinsulin is a single-chain polypeptide comprising both the A- and B-chains of active insulin with a connecting peptide of 30 amino acids residues The connecting or C-peptide apparently serves to place the A- and B-chains in proper position for the establishment of the disulfide bridges When the bridges are deliberately broken in samples of proinsulin, they may be united efficiently without any resultant loss of biological activity The C-peptide is removed intracellularly by the action of a trypsin-like enzyme shortly after the synthesis of proinsulin Insulin is then stored within membrane-bound granules prior to secretion as described above This method of activation by enzymatic digestion of a portion of the parent peptide chain is analogous to trypsin activation of a number of pancreatic digestive enzymes which are secreted as inactive forms

Although most insulin is activated within the β-cell before secretion, about 5% of the

insulin in blood exists as proinsulin Proinsulin has very little biologic activity and there is no evidence that it can be converted to insulin outside the pancreas The discovery of this new step in the biosynthesis of insulin raises the possibility that some forms of diabetes can generate from either genetic alterations in the synthesis of the C-peptide or in the mechanism of activation of insulin

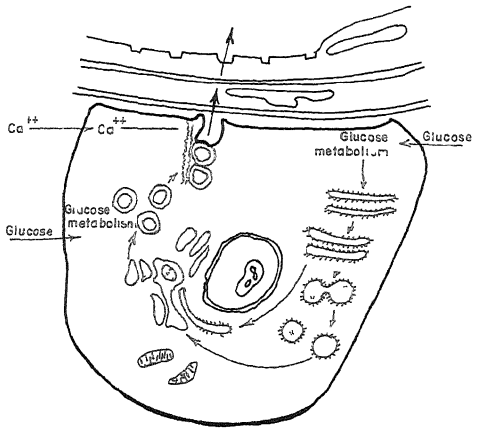


Fig 7 66 Proposed model of β cell secretion Glucose metabolism signals the initiation of proinsulin synthesis in the endoplasmic reticulum Proinsulin is either transferred directly to the Golgi complex or via saccular forms of the endoplasmic reticulum The site of transformation of proinsulin to insulin probably occurs in the Golgi complex and intermediate granules may be present in the sacs of the Golgi apparatus The mature β granules are encased in smooth membranous sacs and discharged into the cytoplasm The granules become associated with the microtubular-microfilamentous system of the β cell calcium influx is produced by glucose metabolism and the calcium ions produce contraction or a change in physical conformation of the microtubular-microfilamentous system with resultant displacement of the granules to the surface of the cell where they are liberated by emiocytosis (From Lacy 1970)

CHEMICAL NATURE AND BIOSYNTHESIS OF GLUCAGON

Glucagon is a single-chain peptide with a molecular weight of about 3500 (Fig 7 68) It is difficult to separate from insulin and is a common contaminant of many commercial preparations of insulin The discovery of glucagon was the result of attempts to understand the nature of a transient hyperglycemia which was noted with many early preparations of insulin Subsequent isolation and characterization of a material, which was known as hyperglycemic factor, revealed a discrete small protein which promoted hepatic glycogenolysis in extremely small concentrations The peptide has 29 amino acid residues and its structure has been confirmed by total chemical synthesis (Wunsch, 1967)

Glucagon is synthesized and secreted by the α-cells of the pancreatic islets Small amounts have also been extracted from gastric and intestinal mucosa of a number of animal species (Marks and Samols, 1970) The nature and properties of gut or enteroglucagon are not well understood, but there is considerable evidence that this latter factor or factors may be of importance in the regulation of insulin se-

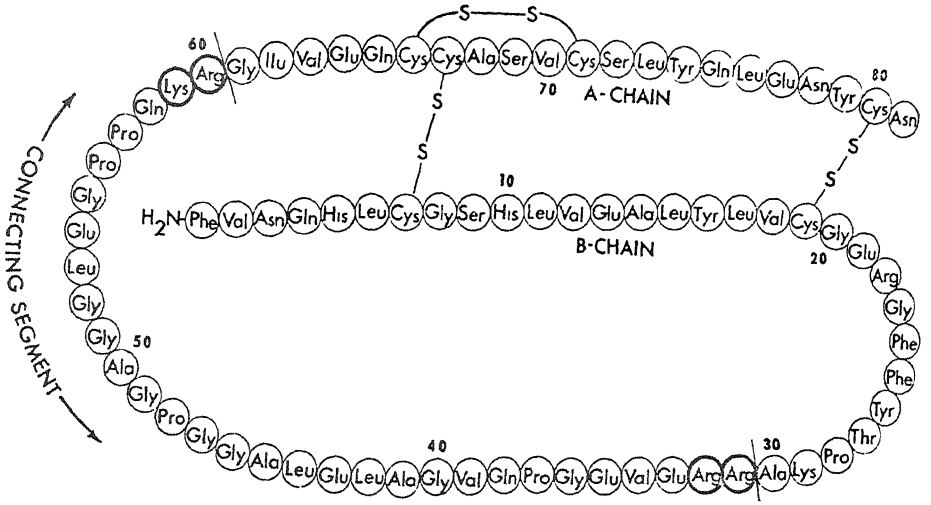


Fig 7 67 Primary structure of beef proinsulin Numbers above sequence indicate position with respect to amino-terminus of the B-chain segment which is amino-terminal in proinsulin (From Nolan et al 1971)

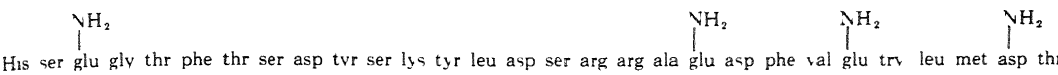


Fig 7 68 Structure of glucagon (From Bromer et al 1957)

cretion Present evidence indicates that there are at least two fractions of immunoreactive glucagon in the gut one has a molecular weight twice that of pancreatic glucagon, another, comprising nearly half of the immunoreactivity, has a molecular size comparable to pancreatic glucagon

MECHANISM OF INSULIN AND GLUCAGON SECRETION (Meyhew et al, 1969)

Insulin (Table 7 7) In a classic study with the perfused rat pancreas, Anderson and Long (1948) clearly demonstrated the primary role of glucose in stimulating insulin secretion Since the response was obtained with the isolated pancreas, it was apparent that an intact nerve supply was not absolutely required Development of improved electron microscopic methods and sensitive immunoassay techniques for the measurement of insulin have greatly stimulated research in this area As a result, a number of additional factors have been shown to be of importance in modulating insulin secretion

As was mentioned in an earlier discussion of islet cells, some investigators have observed striking morphologic changes in β-cell granules following the stimulation of insulin secretion The picture that is presented is one in which insulin granules which are scattered throughout the cytoplasm move toward the outer membrane of the cell The membrane surrounding the granules fuses with the cell membrane, and, at the point of fusion, a hole appears through which the granule is extruded This process, known as emiocytosis, so distorts the cell membrane as to give the appearance of microvilli which are also characteristic of membranes from absorptive cells

With the isolated pancreas, stimulation of insulin secretion by glucose occurs in two distinct phases (Grodsky et al, 1970, Levine, 1970) Immediately after the addition of glucose, insulin is rapidly released in a burst of secretory activity, the rate of secretion then falls With maximal stimulation, only about 2% of the total quantity of insulin present in the pancreas is released With continued application of the glucose stimulus, a lower rate of secretion (Phase II) is maintained for longer periods of time Inhibitors of protein synthesis, such as puromycin, will reduce the second phase and for this reason this phase is thought to be dependent in part, at least, on the release of newly synthesized insulin

Table 7 7
Factors stimulating or inhibiting insulin secretion

Stimulants	Inhibitors
Glucose	2-Deoxyglucose
Mannose	Mannoheptulose
Arginine	Epinephrine
Leucine	
β ketoacids	
Secretin	
Pancreozymin	
Glucagon	
Cyclic AMP	
Theophylline	
Sulfonylureas	

Modified from Levine, 1970

The mechanism of the stimulation of insulin secretion by glucose is not known, but it appears to be closely related, in some way to the utilization of glucose (Randle et al, 1968) Mannose, which can be utilized by the β-cell, will substitute for glucose, but nonmetabolizable sugars such as galactose or D-xylose fail to stimulate secretion Furthermore, interference with glucose metabolism by the addition of mannoheptulose, a seven-carbon sugar, or 2-deoxyglucose, will also block the stimulatory effects of added glucose or mannose From present information, it would appear that the signal generated by glucose for the release of insulin is derived from some intermediate in the pathway of glucose metabolism beyond glucose-6-phosphate

In addition to glucose, amino acids have also been shown to evoke insulin release Ingestion of protein or the infusion of a mixture of amino acids in vivo will cause a substantial increase in levels of plasma insulin which are seemingly unrelated to the slight elevation in glucose that accompanies the infusion The most potent effects are produced by arginine and lysine, although other amino acids, in-

cluding leucine, are effective Christensen (1971) has recently shown that certain amino acid analogues which are transported but not metabolized by the B-cell will also evoke a secretory response The mechanism of the amino acid effect is not understood The ketone bodies, β-OH-butyric and acetoacetic acids, have also been reported to stimulate insulin secretion (Madison et al , 1964)

Cyclic Adenosine Monophosphate (Cyclic AMP). Several lines of evidence strongly implicate cyclic AMP in some fundamental way in the mechanism of secretion Glucagon has been shown to increase cyclic AMP levels in islet tissue and to stimulate insulin synthesis and release from the β-cell (Table 7 8) The stimulatory effect is augmented by agents such as theophylline which inhibit phosphodiesterase, the enzyme responsible for cyclic AMP degradation Moreover, the administration of the cyclic nucleotide itself evokes secretion Analogous effects of cyclic AMP have been noted in the thyroid and the anterior pituitary As in many other secretory cells, the stimulus to secretion appears to be strongly dependent on Ca^{++}, K^{+} is also required At first, investigators were puzzled by the fact that epinephrine either failed to stimulate secretion or exerted an inhibitory effect in the presence of glucose It is now known that epinephrine attaches itself to an α-adrenergic receptor site, which inhibits cyclic AMP formation When an α-adrenergic blocking agent, such as phentolamine, is added, epinephrine exerts its effects via the β-adrenergic site which is stimulatory to cyclic AMP formation and insulin secretion (Table 7 8)

In addition to the effects of glucagon and epinephrine, the intestinal hormones, secretin and pancreozymin, are known to stimulate insulin secretion (Marks and Samols, 1970) The role of these hormones in modulating the insulinotropic effects of glucose and glucagon will be discussed below Tolbutamide, a sulfonylurea, which has been used in the treatment of maturity-onset diabetes, is an effective stimulus of insulin secretion This agent apparently works independent of the stimulation by glucose since mannoheptulose, which blocks the effect of glucose on insulin release, fails to inhibit the tolbutamide effect

Glucagon Secretion (Marks and Samols, 1970) Owing to the presence of interfering and cross-reacting materials in plasma, glucagon has been more difficult to measure by immunologic techniques than has insulin For this reason, comparatively little information is available either on the mechanism of secretion or the nature of agents that stimulate or inhibit release Severe hypoglycemia that is induced by phloridzin in dogs has been shown to promote glucagon secretion However, glucagon secretion is either unaffected or decreased somewhat with prolonged fasting While high glucose concentrations in the blood perfusing the pancreas will apparently suppress glucagon secretion, moderate degrees of hypoglycemia are seemingly not a stimulus to glucagon secretion Fasting may indirectly promote secretion by other mechanisms Both pancreozymin and amino acids have been shown to exert potent stimulatory effects in vitro

Secretion under Physiological Conditions. There is general agreement that glucose is the

Table 7 8

Effects of theophylline glucagon epinephrine α-adrenergic blocking agent (phentolamine) and β-adrenergic blocking agent (propranolol) on pancreatic islet cyclic 3′5′ AMP content and insulin release in the rat

Experimental Conditions	Theophylline	Cyclic 3′5′ AMP Concentration in Pancreatic Islets	Insulin Secretion in Vivo
		μmoles/g	*μunits-minutes ml* $^{-}$
Control	−	355 ± 50	0
	+	5,570 ± 302	487 ± 42
Epinephrine	+	415 ± 12	208 ± 29
Epinephrine + phentolamine	+	4,450 ± 363	1,272 ± 36
Epinephrine + propranolol	+	305 ± 20	0
Glucagon	+	11,950 ± 431	1,023 ± 97

From Turtle and Kipnis, 1967

primary agent which regulates insulin secretion Plasma insulin levels increase promptly in response to hyperglycemia, and secretion is diminished during hypoglycemia and fasting It would appear that epinephrine, which may be released in response to hypoglycemia, is able to exert a restraining effect on insulin secretion

It has long been observed that the magnitude of the hyperglycemia following oral glucose is appreciably smaller than that resulting from a comparable quantity of glucose administered intravenously (Fig 7 69) This difference has been attributed to an augmentation of insulin secretion by intestinal factors which are mobilized during the absorption of glucose in the gut These factors have not been identified but intestinal glucagon is a possibility Both secretin and pancreozymin are known to stimulate insulin secretion, but it is unlikely that the elaboration of these intestinal hormones is significantly increased by the absorption of glucose Other foodstuffs, such as amino acids and fat, will evoke pancreozymin secretion, which in turn may stimulate the release of insulin Although this sequence would not explain the increased release of insulin in response to the absorption of glucose, it does suggest a possible mechanism whereby the process of digestion can promote insulin secretion in anticipation of increased blood levels of substrates

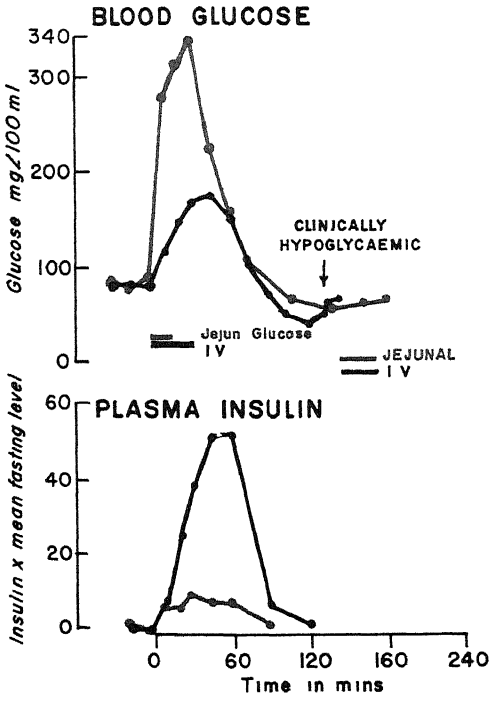


Fig 7 69 Blood glucose and plasma insulin response to rapid infusion of 60 g of glucose either intravenously or intrajejunally to a normal subject (From McIntyre et al, 1965)

FATE OF INSULIN AND GLUCAGON AND REACTION WITH SYSTEMIC TISSUES

Measurements of plasma insulin turnover by disappearance of ^{131}I insulin or immunoreactive insulin has shown that insulin is rapidly removed from plasma by the body tissues (Williams, 1968) The half-life of plasma insulin is 10 min or less which makes it necessary for the pancreas to secrete insulin steadily in order to maintain levels that are sufficient for optimal metabolic performance Little or no insulin is taken up by the brain or red cells, but the liver, kidneys, and skeletal muscle remove almost all of the insulin that is secreted Very little is excreted in the urine The relative affinity of the tissues for insulin appears to bear no relationship to the biological effects of the hormone on the tissues, instead, the very large uptape of circulating insulin by the liver and kidneys is thought to be primarily related to the need of the body to degrade insulin This function facilitates regulation of plasma insulin by providing a means to decrease insulin concentrations when insulin secretion is cut off Although skeletal muscle is one of the important sites of insulin action, the removal of plasma insulin per unit mass of tissue is relatively small as compared with either the kidney or the liver However, owing to the large mass of muscle, the total quantity of insulin removed is similar to that degraded by the kidneys or the liver

Knowledge of the nature and location of insulin-binding sites on cells is of great importance to further understanding of the mechanism of insulin action In general, the binding of insulin appears to serve at least two functions The first involves the attachment of the hormone to receptor sites which directly or indirectly alter the rates of metabolic functions The nature of the metabolic effect depends on the particular tissue involved, as well as its metabolic state. It is clear from recent studies by Cuatrecasas (1969) with isolated fat cells that most of the well-known effects of insulin in that tissue can be mediated from

receptor sites on the cell surface In these particular experiments, the insulin that was used was chemically attached by the N-terminal amino acid of the B-chain to a very large agarose polymer particle The modified insulin was reacted with isolated fat cells and found to retain all of the biological activity of native insulin Since the large size of the agarose prevented penetration of insulin into the interior of the cell, it appeared to act through a receptor site that was on or close to the outer membrane of the cell While this experiment does not eliminate the possibility that there are other receptor sites within cells, it is in keeping with a large body of evidence indicating that many of the metabolic effects, including those concerned with membrane transport, are mediated at sites on the cell membrane

The second function served by insulin binding is the destruction of insulin As mentioned above, the necessity for such a mechanism becomes clear when one considers that metabolic regulation is a bidirectional process requiring downward as well as upward adjustments in hormone levels The continuous secretion of insulin by the pancreas, which in the human averages 1 to 2 mg /day, is determined in part by the rate of destruction of insulin in peripheral tissues All tissues that respond to insulin appear to be capable of degrading insulin to some extent The converse is not true, however, since the kidney, which is apparently unaffected by insulin, does bind and degrade insulin to a considerable extent The process of insulin degradation has been studied extensively in liver, and studies employing the isolated perfused liver have shown that the capacity for destroying insulin far exceeds the limits of the tissue to respond to the metabolic effects of the hormone This indicates that sites of insulin inactivation may differ from those concerned with metabolic regulation Little is known of the insulin-inactivating mechanism except that the hormone is known to be completely broken down to its constituent amino acids An enzyme, insulin-glutathione transhydrogenase, in liver has been found to catalyze the reductive cleavage of the disulfide bridges in insulin (Tomizawa, 1962, Katzen et al., 1963) For a time it was thought that this enzyme played an important role in insulin inactivation Results of more recent studies, however, make it doubtful that this enzyme plays a specific role in insulin destruction

The question whether binding of insulin involves a covalent bonding between the hormone and reactive groups on the receptor site has been raised One possibility would be the breaking of the disulfide bridges and reattachment of insulin to free SH-groups on the receptor sites Preliminary attempts to study this possibility have indicated that bonding of this type may occur but the evidence is far from conclusive The well-known fact that the metabolic effects of insulin can be promptly stopped by the addition of insulin-binding antibody indicates that the attachment mechanism is readily reversible

Systemic Effects of Insulin and Glucagon

INSULIN

Insulin exerts a variety of metabolic effects on many tissues of the body but its fundamental mode of action or actions at the cellular level is unknown Hormonally-responsive tissues include muscle (skeletal and heart), adipose tissue, liver, white cells, mammary glands, bone, skin, lens, pituitary, peripheral nerve, and aorta Brain, gonads, and red cells are notable for their lack of responsiveness to insulin, the question as to whether the kidney is responsive is a matter of debate The failure of insulin to cross the blood-brain barrier may explain the lack of responsiveness by the brain, but there is no explanation for the other insulin-insensitive tissues except for the apparent absence of specific receptor sites Since metabolic responses to insulin differ among the various tissues according to the functional differentiation of the cells, the role of insulin in regulating specific metabolic processes will be discussed in chapters dealing with these processes

One of the primary roles of insulin in the overall economy of the body is to regulate uptake, storage, and release of important circulating substrates—glucose, amino acids, and fatty acids The levels of these constituents in plasma reflect the sum of their individual rates of production and utilization by the body tissues Many of these processes are strongly influenced by insulin and glucagon (see chapt 10 through 13) Inasmuch as the secretion of in-

sulin and glucagon is in turn affected by the level of circulating glucose and amino acids (and possibly ketone bodies), the dominant pattern of regulation appears to be of the feedback type This is now well-documented in the case of regulation of plasma levels of glucose and a similar pattern for regulation of plasma amino acid levels is now beginning to emerge

Glucose Regulation. Blood glucose is maintained within reasonably narrow limits under a wide variety of metabolic conditions and, even with prolonged fasting, values seldom fall below 60 mg /100 ml of blood This regulation requires the coordinated control of metabolic processes in many tissues in order to balance glucose utilization against glucose production Many tissues, such as skeletal muscle, brain, heart, and adipose tissue consume sizeable quantities of glucose, but only the liver is capable of releasing sufficient amounts into the circulation during the postabsorptive period to balance the rates of utilization The vital role of the liver as a source of glucose was shown several years ago by Mann and Magath (1924) who studied the effects of hepatectomy in dog Removal of the liver resulted in hypoglycemia which invariably proves fatal unless glucose is administered Although glucose is stored in the form of glycogen in muscle, absence of glucose-6-phosphatase prevents its release as free glucose into blood The solution to this impasse is conversion of glucose to lactate by glycolysis in muscle, but in the absence of liver, there is little reconversion of the lactate to glucose, a process known as gluconeogenesis While the kidney possesses the enzymes that are required for gluconeogenesis, the quantitative capacity is not sufficient to maintain normal glucose levels in the absence of liver

The two major pathways of glucose formation in liver, viz , glycogenolysis and gluconeogenesis, are regulated by insulin When rates of glucose formation are increased by agents such as glucagon or epinephrine, which increase tissue levels of cyclic AMP, the addition of insulin exerts a strong inhibitory effect on glucose formation and release from the liver, presumably by its effect to reduce tissue levels of cyclic AMP (chap 12) In addition to these hormonal factors, blood glucose itself strongly inhibits hepatic glucose release and stimulates its uptake by effects which are exerted predominantly on glycogen metabolism Attempts to demonstrate effects of insulin on hepatic glucose production in vivo have been only partially successful due to the fact that an effect of insulin to reduce glucose output may be masked by an opposite effect of the hypoglycemia to increase glucose output (deBodo et al , 1963) Since the effect of insulin to lower blood glucose can be attributed to stimulation of glucose uptake by muscle and adipose tissue, the role of the liver in glucose regulation was not readily recognized Direct measurements of hepatic glucose output during insulin hypoglycemia generally have shown that output remains unchanged If, however, glucose is administered with insulin in amounts sufficient to prevent the hypoglycemia, hepatic glucose output is significantly reduced One may deduce from these findings that the failure of the liver to alter glucose release during insulin-induced hypoglycemia represents a very strong effect of insulin to restrain glucose output By contrast, the effect of agents such as phloridzin, which can reduce blood glucose by blocking the reabsorption of glucose in the renal tubule, evokes a large increase in hepatic glucose output An acute deficiency of insulin produced in the rat by administration of anti-insulin serum or mannoheptulose is associated with an immediate increase in hepatic glucose release and a rise in blood glucose which reaches a new steady state In this situation, reduction in circulating insulin will not reduce the actual uptake of glucose by muscle and other peripheral tissues since the level of glucose also increases Thus, the upward or downward adjustment in the level of insulin results in significant changes in the steady-state levels of blood glucose with only temporary alterations in glucose utilization and production

In the human, as opposed to the rat liver, glycogen is comparatively stable and is not depleted readily for glucose production during the postabsorptive period Most of the glucose that is produced by liver under these conditions is derived by gluconeogenesis (see chap 12) Approximately 20% of the glucose which is produced in this way is synthesized from lactate derived by glycolysis in the red cell, skeletal muscle, and other tissues This fraction may be thought of as a pool of carbon which is continuously recycled through the glycolytic and gluconeogenic pathways However, a major fraction of the glucose that is released is oxidized to CO_2 or utilized for lipogenesis and

other synthetic reactions Since this carbon cannot reappear as glucose, it must be replaced by carbon from other sources Amino acids from the breakdown of tissue protein contributes the largest fraction of carbon for gluconeogenesis, glycerol derived from the hydrolysis of esterified fatty acids also makes a significant contribution During the immediate postabsorptive period and during short periods of fasting, the brain is known to subsist entirely on glucose as a source of fuel By contrast, such tissues as muscle oxidize fatty acids and ketone bodies in preference to glucose Thus, there appears to be an absolute requirement for new glucose synthesis based on the metabolic needs of the central nervous system During prolonged fasting, the brain can oxidize ketone bodies that are derived from the breakdown of fatty acids (Cahill et al , 1970) This slower adaptive response provides an additional safeguard against the ultimate depletion of gluconeogenic reserve

The necessity for restocking carbon stores presents itself at intervals and is signaled by the awareness of hunger The metabolic events leading to the repletion of stores are initiated by the ingestion of food Insulin secretion increases in response to the rise in blood glucose and possibly also to the increased elaboration of intestinal hormones, such as enteroglucagon and pancreozymin With increased rates of insulin secretion and absorption of glucose and amino acids, levels of glycogen and protein increase to replace deficits that were incurred during the previous postabsorptive period The largest changes in carbon stores per unit of organ weight occur in liver, corresponding changes in skeletal muscle are far smaller by comparison At the same time, fat stores in adipose tissue are repleted The extent to which the glycogen deposition is the result of diminished rates of glycogenolysis as opposed to the stimulation of glycogen synthesis is not established with certainty, but it is likely that both mechanisms operate Insulin has been shown to stimulate glycogen synthesis in liver of alloxan diabetic rats (Steiner and King, 1964) Recent studies indicate that acetylcholine or electrical stimulation of the vagus nerve may also potentiate glycogen synthesis in rat liver and that this cholinergic effect may be of importance in the regeneration of fuel stores during feeding

Amino Acids. The role of insulin in the regulation of blood levels of amino acids is not as well-documented as in the case of glucose Certain features, however, are now emerging which suggest that a similar type of control exists It has long been known that the injection of insulin will promptly decrease blood amino acid levels in vivo Such an effect could be mediated by one or more of the following possible responses to insulin which have been demonstrated experimentally in isolated tissues *1*) increased membrane transport of certain amino acids in muscle, 2) increased incorporation of amino acids into muscle, and possibly liver protein, and *3*) decreased proteolysis in liver and muscle Amino acids as well as insulin appear to exert any inhibitory effect on proteolysis in the liver and muscle, in addition to potentiating protein synthesis In studies with the isolated perfused liver, net changes in liver protein breakdown can be followed by the release of valine since the latter is not metabolized significantly by liver Some inhibitory effects of infused insulin and amino acids on net valine release under these conditions are shown in Figure 7 70 Since amino acids have been shown to stimulate insulin secretion in the isolated pancreas or in vivo

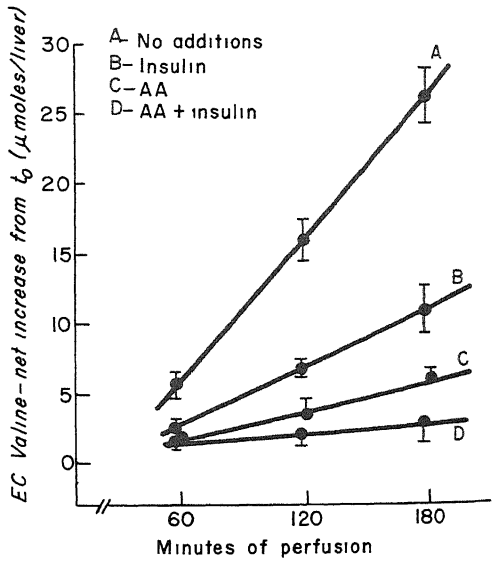


Fig 7 70 Effects of insulin and amino acid infusions on the accumulation of extracellular valine during perfusion of livers from nonfasted rats weighing about 125 g Insulin and an amino acid mixture containing glutamic acid lysine histidine arginine aspartic acid glycine proline, alanine serine cysteine methionine threonine phenylalanine and tryptophan were infused as indicated (From Mortimore 1971)

when administered either orally or intravenously, there is a strong possibility that blood amino acid levels are regulated by a feedback mechanism analogous to that previously described for glucose Much work remains to be done before such a possibility is established, but the need for amino acid regulation is clearly evident from the point of view of providing steady-state concentrations of free amino acids for protein synthesis and gluconeogenesis during the postabsorptive state

GLUCAGON

The action of glucagon on the regulation of circulating substrates is limited mainly to the liver where it is known to exert strong effects to promote glycogenolysis (and to inhibit glycogen synthesis), gluconeogenesis, lipolysis, and ureogenesis We have mentioned earlier its potentiating effect on insulin secretion by the β-cell Glucagon is known to stimulate lipolysis in adipose tissue, but the physiological importance of this effect is still a matter of debate (see chap 11) As far as it is known, glucagon exerts no effects in skeletal muscle, although it is capable of promoting glycogenolysis in the heart where it also exerts a strong inotropic effect

In liver all of the known effects of glucagon can be attributed to its primary action in stimulating the biosynthesis of cyclic AMP Under the influence of glucagon, processes concerned with glucose formation are accelerated, thus increasing rates of glucose output by the liver In some avian and amphibian species, glucagon appears to be a dominant factor in glucose regulation and pancreatectomy in these animals results in hypoglycemia rather than hyperglycemia In mammals, however, the stimulatory effects of glucagon (and catecholamines) are normally suppressed by insulin Studies employing the isolated perfused rat liver, a preparation which is devoid of most extrinsic influences, indicate that some continuous stimulation of glucose release by glucagon or catecholamines is required for insulin to exert its inhibitory effects (Fig 7 71) The present view is consistent with the notion that insulin (and glucose) are responsible for much of the moment-to-moment regulation of glucose output by the liver, the magnitude of this regulation, however, is set by a more "tonic" influence of glucagon or other agents to maintain adequate levels of tissue cyclic AMP This

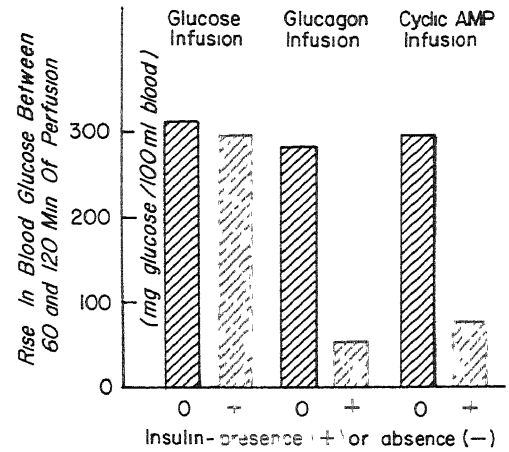


Fig 7 71 Effect of insulin on hepatic glucose production Livers were obtained from nonfasted rats All additions were made by infusion into the medium after a 60-min period of equilibration (Modified from Glinsmann and Mortimore 1968)

notion is also compatible with what is presently known of the regulation of insulin and glucagon secretion Thus, while insulin secretion is sensitive to small fluctuations in blood glucose, glucagon levels are comparatively stable and are affected only by more extensive deviations in glucose concentration

Glucagon is also a potent stimulator of ureogenesis, an effect which has been demonstrated both in vivo and in vitro The effect is closely linked with the stimulation of gluconeogenesis and increased amino acid oxidation (chap 12) Net protein breakdown in liver is also increased by glucagon but it is not known whether the effect is primarily to stimulate proteolysis or to inhibit protein synthesis With regard to the interaction between insulin and glucagon on regulation of amino acid metabolism, it is clear that, in contrast to the regulation of glucose, insulin is able to inhibit the release of amino acids from liver in vitro in the absence of any added glucagon

Physiological Aspects of Insulin Deficiency

INSULIN ANTAGONISTS

There are a number of factors of physiological importance whose actions, in general, are opposite to those of insulin One of these, glucagon, has been discussed, the others, epinephrine, glucocorticoids, growth hormone, and

fatty acids, are discussed below Normally, the effects of these agents are well-integrated with those of insulin However, when one or more of these antagonists is greatly increased or decreased in amount, then the effectiveness of insulin will be decreased or increased accordingly, giving rise to significant alterations in metabolic regulation

Epinephrine. Epinephrine, by its action to increase cyclic AMP, is similar to glucagon Its effects are more widespread, however, and in addition to liver, tissues such as muscle and adipose tissue are highly responsive Of all the known physiological antagonists, it is perhaps the most rapidly acting In fight-or-flight responses, the release of epinephrine or related catecholamines into plasma or at nerve-endings promotes rapid glycogenolysis in liver and muscle and stimulates lipolysis in adipose tissue and gluconeogenesis in liver Owing to the increased levels of fatty acids in plasma, glucose utilization in muscle and other peripheral tissues will be reduced Blood glucose increases rapidly, presumably to ensure adequate energy-yielding substrate for use by the central nervous system, the rise in free fatty acids provides an abundant source of fuel for oxidative processes in muscle

Glucocorticoids. Adrenal cortical steroids, such as cortisol, are potent catabolic agents which exert widespread effects in nearly all tissues of the body They play important roles in promoting lipolysis and gluconeogenesis and in the latter process, for example, steroids are responsible for the maintenance of much of the enzyme machinery in the liver In fact, the stimulating effect of glucagon on gluconeogenesis cannot be expressed in the absence of glucocorticoids (chap 12) In adrenal cortical insufficiency, the ability of the liver to mobilize glucose during fasting is limited, a situation which may lead to fatal hypoglycemia The opposite extreme is seen in instances of chronic glucocorticoid excess, where excessive rates of proteolysis and gluconeogenesis lead to nitrogen depletion Since insulin is relatively ineffective in restraining these alterations, overproduction of glucose coupled with its decreased utilization in systemic tissues gives rise to moderate hyperglycemia The pancreatic islets will be stimulated to secrete insulin in excessive amounts In patients with early diabetes the secretory capacity of the islets is limited and severe clinical diabetes may result

Growth Hormone In addition to anabolic activity, growth hormone has long been known to be diabetogenic (chap 13) The acute metabolic effects of growth hormone simulate those of insulin, but they are transient and the continued administration of this agent, experimentally, has been shown to produce hyperglycemia, degenerative changes in the β-cells of the pancreatic islets, and eventually permanent diabetes in some mammalian species The initial hyperglycemia is associated with striking increases in plasma insulin, a fact which suggests that part of the diabetogenic mechanism of growth hormone may involve the antagonism of insulin at effector sites in peripheral tissues, a situation which could bring about a chronic overproduction of insulin and eventual exhaustion of the β-cell There is also evidence that growth hormone may stimulate hypersecretion of the β-cell directly The marked amelioration of diabetes produced by hypophysectomy that was first described by Houssay and Biasotti (1931) can be attributed in part to the loss of this growth hormone effect The loss of adrenocorticotrophic hormone (ACTH) and diminished secretion of glucocorticoids also contribute to the amelioration

Effect of Fatty Acids. Fatty acids and ketone bodies are known to be utilized by skeletal muscle and heart in preference to glucose as an oxidizable substrate (chap 10) In situations where lipolysis is accelerated, fatty acid levels in plasma are increased and rates of glucose phosphorylation in muscle will decrease Thus, the effectiveness of insulin in regulating glucose utilization will be diminished This fatty acid effect is thought to account in part for the glucose intolerance and insulin resistance seen in fasting and certain forms of diabetes where the breakdown of fat stores is a dominant feature Whether or not it accounts for the insulin resistance seen in obesity has not been established

Antagonism by Insulin-Binding Antibodies. Insulin resistance from the development of insulin-binding antibodies is not infrequently encountered in diabetic patients taking insulin Antibodies develop in response to the injection of foreign protein, in this case from another mammalian species, and the effect is to inactive large quantities of the in-

jected insulin Treatment involves changing to an insulin preparation from an animal source which does not cross-react immunologically

DIABETES MELLITUS
(Camerini-Davalos and Cole 1970)

In the majority of instances the cause of this disorder is a genetic defect in insulin secretion by the β-cell The expression of this defect ranges from minimal in the case of pre- or early diabetes to an absolute insulin deficiency seen in severe juvenile diabetes In most cases of clinically recognized diabetes, the defect falls somewhere between the two extremes and the metabolic consequences are of moderate severity Since under normal conditions the amount of insulin secreted by the β-cell represents only a small fraction of its maximal secretory capacity, diabetes is not ordinarily manifested until most of the secretory reserve is lost It should be emphasized, however, that when insulin antagonists are present, as with stress-induced glucocorticoid secretion, the need for insulin may exceed the existing secretory capacity and the metabolic consequences of severe insulin lack will often appear

The question has been raised as to whether certain forms of maturity-onset diabetes, often associated with obesity, are primarily the result of an acquired resistance to insulin since in some instances plasma insulin values have been found to be increased and the insulin response to oral glucose exaggerated More recent studies, however, indicate that the increase in insulin resistance associated with obesity is not directly linked with an inherent defect in the ability of the B-cell to secrete insulin Thus, while the insulin resistance associated with obesity may act like growth hormone or fatty acids in antagonizing the effects of insulin, the phenomenon may or may not be associated with a genetic defect in insulin secretion

EFFECTS OF ABSOLUTE INSULIN DEFICIENCY— KETOACIDOSIS

With the complete loss of insulin effectiveness, a series of metabolic derangements takes place which eventually lead to ketoacidosis Since the forces which stimulate lipolysis and gluconeogenesis are unrestrained in the absence of insulin, the rates of these processes will rapidly increase, resulting in high levels of glucose and free fatty acids in plasma The breakdown of protein in liver and muscle will increase, thus providing a steady source of carbon for gluconeogenesis Since the quantity of free fatty acids presented to the liver will greatly exceed and capacity of liver to oxidize these substrates to CO_2, fatty acid oxidation will proceed only through β-oxidation and the partially oxidized products will accumulate as ketone bodies, β-OH-butyric acid, acetoacetic acid, and acetone These will spill into the blood where the concentrations may reach values as high as 90 mg /100 ml of plasma Since β-OH-butyric and acetoacetic acids are moderately strong, blood pH will drop and a state of metabolic acidosis will ensue Owing to the extreme hyperglycemia which may reach values as high as 400 to 800 mg /100 ml of blood, glycosuria is pronounced and will continue as long as there is sufficient gluconeogenic reserve If the metabolic defect is not corrected by the administration of insulin, further complications, largely of renal origin, will appear Owing to the excretion of large volumes of water which accompanies the glucose, dehydration develops and with the loss of fixed base in the form of Na^+ and K^+, the acidosis becomes even more severe In an attempt to compensate for the acidosis, respiration becomes rapid and deep (Kussmaul breathing) and the kidney, in an effort to conserve the fixed base, will excrete ammonia With further dehydration, decreased plasma volume, peripheral circulatory failure, and coma terminated by death are the usual sequence of events

Members of our society are taught from early years to distinguish men from women, however, sex identification is a comprehensive concept implying more than gender recognition Sexual differentiation begins at the time of fertilization and continues into early adult life This is a complex process which is influenced independently by genetic, hormonal, and behavioral determinants When we speak of the sex of an individual we mean the composite of the chromosomal sex, gonadal sex, genital sex, and psychological sex In this chapter the development and interrelation of each of these parameters are discussed

GENETIC CHROMOSOMAL, OR NUCLEAR SEX

Genetic sex is determined at the time of fertilization when the spermatozoon and ovum unite and depends upon which sex chromosome is present in the sperm cell All extragonadal cells in the human contain 23 pairs of chromosomes (22 pairs of autosomes and 1 pair of sex chromosomes) In the male the sex chromosomes are X and Y, whereas in the female there are two X chromosomes During the process of ova and sperm development germ cells divide by a process called myosis so that pairs of chromosomes are split between two daughter cells As a result each gamete contains 22 autosomes and 1 sex chromosome In the testis myosis produces sperm which contain X and others containing Y chromosomes Ova, however, contain only X chromosomes At the time of fertilization the sex of the resulting zygote will be determined by the sex chromosome carried by the fertilizing sperm If an ovum is fertilized by a sperm bearing an X chromosome, then the combination XX results and a female develops On the other hand, if the fertilizing sperm bears a Y chromosome, then the XY combination produces a male

In recent years techniques have been developed for examining human chromosomes Cells in metaphase are spread and the chromosomes are stained for counting or photography and mounting A systematized array of chromosomes from a single cell is called a karyotype (Figs 7 72 and 7 73) The meaning of this term is usually extended to imply that the chromosomal pattern of the cells examined are typical of all the cells of the individual To accurately determine the karyotype, however, one must characterize the chromosomes from several tissues and at least 100 metaphase cells from each tissue should be counted The karyotype of a normal woman designated 46, XX (Fig 7 72) and that of a normal male 46, XY (Fig 7 73)

Biologic function of the sex chromosomes has been inferred from the examination of patients with an abnormal number of sex chromosomes In individuals with an XXY (Ford et al , 1959), or an XXXXY (Barr et al , 1963) complement of sex chromosomes, testes develop indicating that the Y chromosome carries strong male-determining genes which can induce fetal testicular development even in the presence of two or more X chromosomes Without a Y chromosome testicular development does not usually occur Only a few genes which influence extragonadal structures have been identified on the Y chromosome

The biologic functions of the X chromosomes are more diverse (Moore, 1966) The requirement for a pair of X chromosomes for ovarian differentiation is suggested by the fact that ovaries do not develop in women with an unpaired X chromosome (Turner's Syndrome, 45, XO) In contrast to the Y chromosome, which is mainly concerned with sex differentiation, the X chromosome contains genes which influence the differentiation of a wide variety of extragonadal characteristics

Sex Chromatin (Nuclear Sex). Since the chromosomal composition of the male differs from that of the female, it was not surprising that Barr and Bertram (1949) observed a mor-

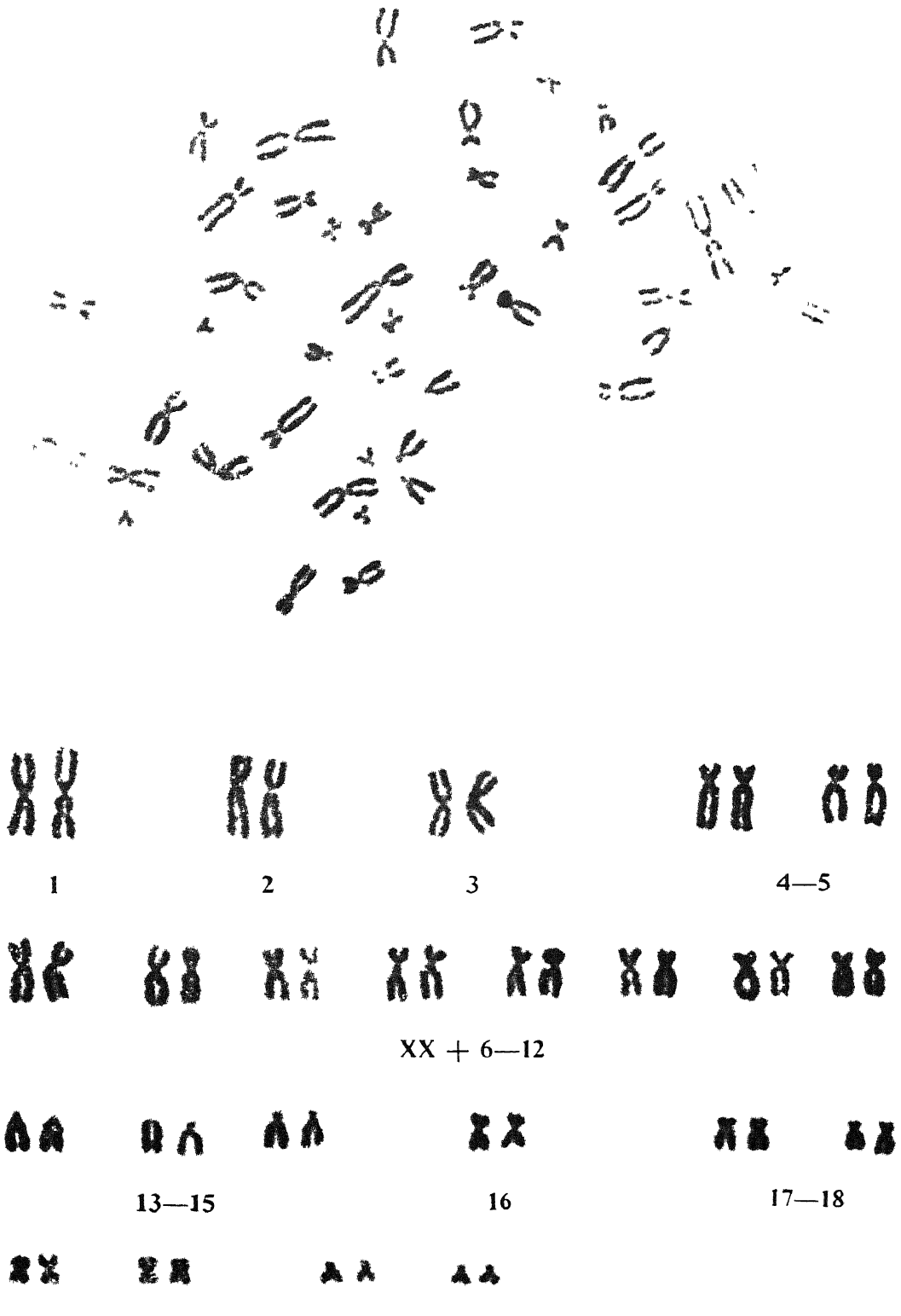


Fig 7 72 Karyotype of blood lymphocytes from a woman The pair of X chromostmes is placed at the left hand of the second row

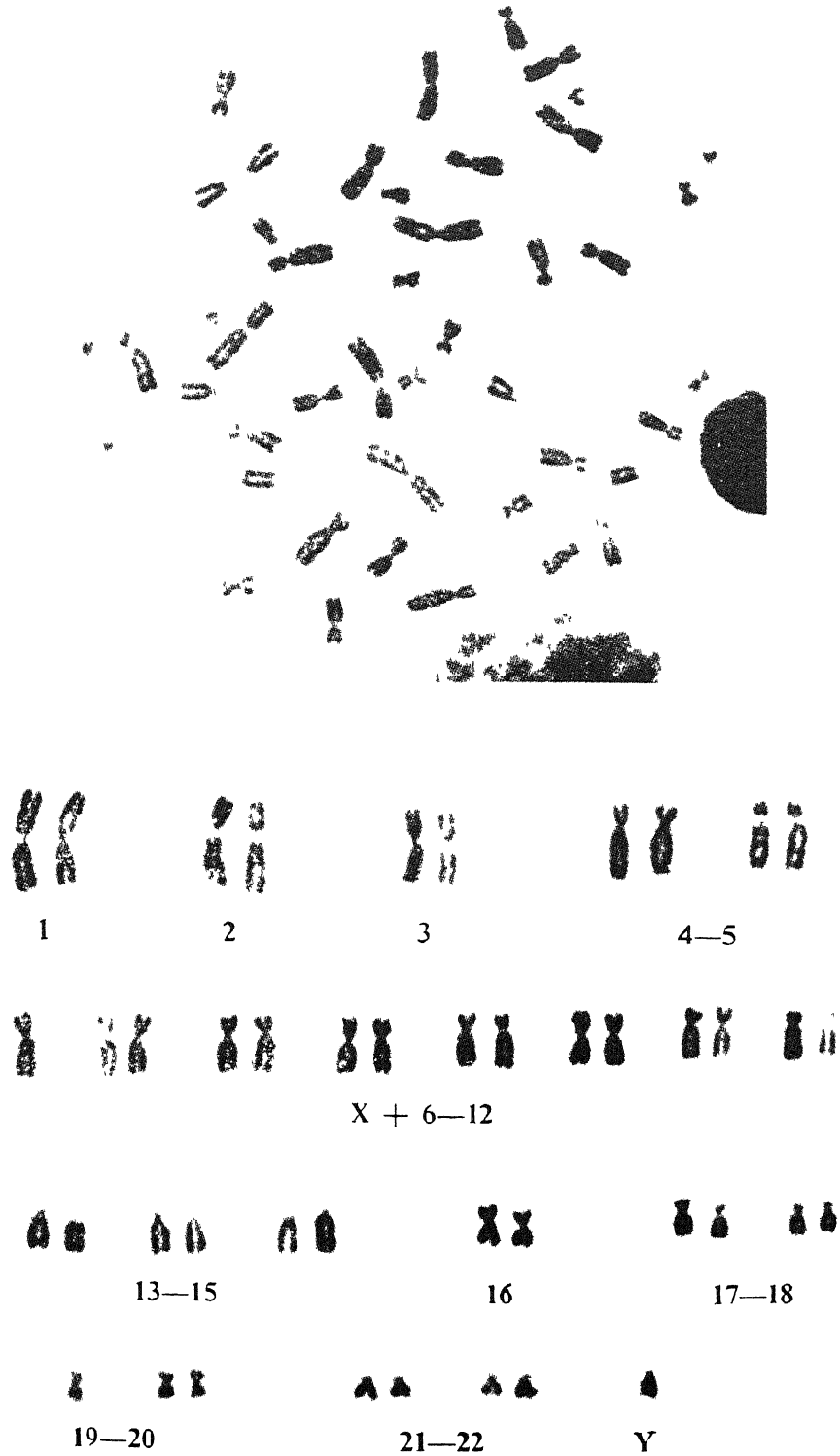


Fig 7 73 Karyotype of blood lymphocytes for a man The X chromosome is at the left of the second row and the Y chromosome is at the bottom right

phologic difference in nuclei from male and female animals The difference between the male and the female lies in a special mass of chromatin (sex chromatin or Barr body) in the female The sex chromatin is typically a plano-convex mass about 1 μ in diameter which stains positively for deoxyribonucleic acid (DNA) and which lies close or in contact with the nuclear membrane (Fig 7 74) It is now generally accepted that the sex chromatin is comprised of one of the X chromosomes which remains genetically inactive during interphase Chromatin positive cells which contain two X chromosomes (46, XX, 47, XXY) have a single Barr body, whereas those with more than two X chromosomes may have two chromatin bodies It follows, therefore, that the number of chromatin bodies in any diploid nucleus is one less than the total number of X chromosomes (Mittwoch, 1964)

GONADAL SEX

Until the 5th or 6th week of embryonic life the gonad in both sexes develops in an identical fashion with the potential to develop into either a testis or an ovary (Grumbach and Barr, 1958) At the 6th week the embryonic gonads appear as ridges on either side of the dorsal mesentery immediately adjacent to the developing adrenal gland The outer layer or cortex is comprised of coelomic epithelial cells into which the primary germ cells migrate from the endoderm of the yoke sac The inner or medullary portion of the indifferent gonad is comprised of stromal mesenchyme which surrounds cords of epithelial cells (primary sex cords) which are derived from the overlying coelomic endothelium If the indifferent gonad is to become a testis due to the presence of Y chromosomes a marked proliferation of the primary sex cords occurs in the medulla about the 7th week of fetal life Primary germ cells migrate from the cortical area and are incorporated into the developing seminiferous tubules About the 8th week Leydig cells appear in the medullary mesenchyme and their function assures subsequent male differentiation of the genital ducts and external genitalia As testicular differentiation proceeds the medulla of the testis progressively enlarges and the cortex regresses In the mature testis the only remnant of the embryonic cortex is the tunica vaginalis

Differentiation of the indifferent gonad into an ovary occurs at about 9 weeks of fetal life It is of interest to note that primary germ cells of a fetus with a 46, XX karyotype have a sex chromatin body (inactive X chromosome) prior to their migration into a genital ridge Once inside the embryonic gonad, however, the sex chromatin body is lost, suggesting that both X chromosomes in the primordial germ cell are active (Witschi, 1957) As noted above, *two* active X chromosomes are believed to be necessary for differentiation of an ovary In the developing ovary the embryonic cortex proliferates to form the major portion of this gonad The medulla regresses and forms the hilum of the mature ovary

From these observations it is apparent that the gonadal sex is to a large extent determined by the chromosomal sex and in most instances both are the same By contrast, genital sex and secondary sexual characteristics are hormonally determined and can easily be made inconsistent with the genetic sex by treatment with sex steroids

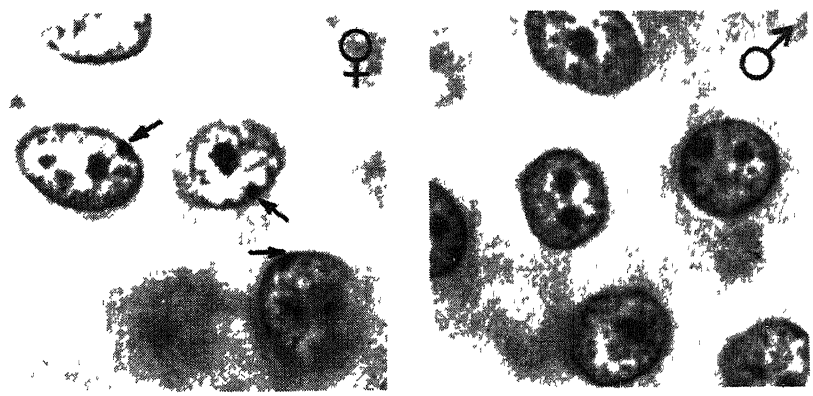


Fig 7 74 Nuclei of cells in the epidermal spinus layer of a genetic female (left) and a genetic male (right) The sex chromatin bodies are indicated by arrows in the female (From Grumbach and Barr 1958)

GENITAL SEX

Differentiation of the Genital Ducts. At the 7th week of embryonic life the fetus contains primordia of the genital duct systems for both the male (Wolffian ducts) and the female (Mullerian ducts) If a male develops, the Wolffian duct system will give rise to the epididymis, vas deferens, seminal vesicles, and the ejaculatory duct, whereas, in the female the Mullerian ducts serve as the anlage of the Fallopian tubes and uterus During the 3rd month of fetal life either the Mullerian or the Wolffian ducts complete their differentiation while the others regress

The mechanism whereby one set of ducts dominates over the other to form the internal genitalia of the developing embryo was demonstrated in a series of experiments by Jost (1953) In these studies embryonic rabbits were either gonadectomized or treated with steroids in utero and the subsequent effect upon Mullerian and Wolffian development observed The results of these studies are schematically summarized in Figure 7 75 In the presence of a testes Wolffian structures develop and Mullerian structures regress (Fig 7 75*A*, *B*, *C*) The development of normal Mullerian structures is not dependent upon the ovary since this ductal system differentiates in both the castrate female and male even if the castration is unilateral (Fig 7 75*D*, *E*) Transplantation of a testis into the female adjacent to the ovary results in unilateral Wolffian differentiation associated with Mullerian regression on the side of the trans-

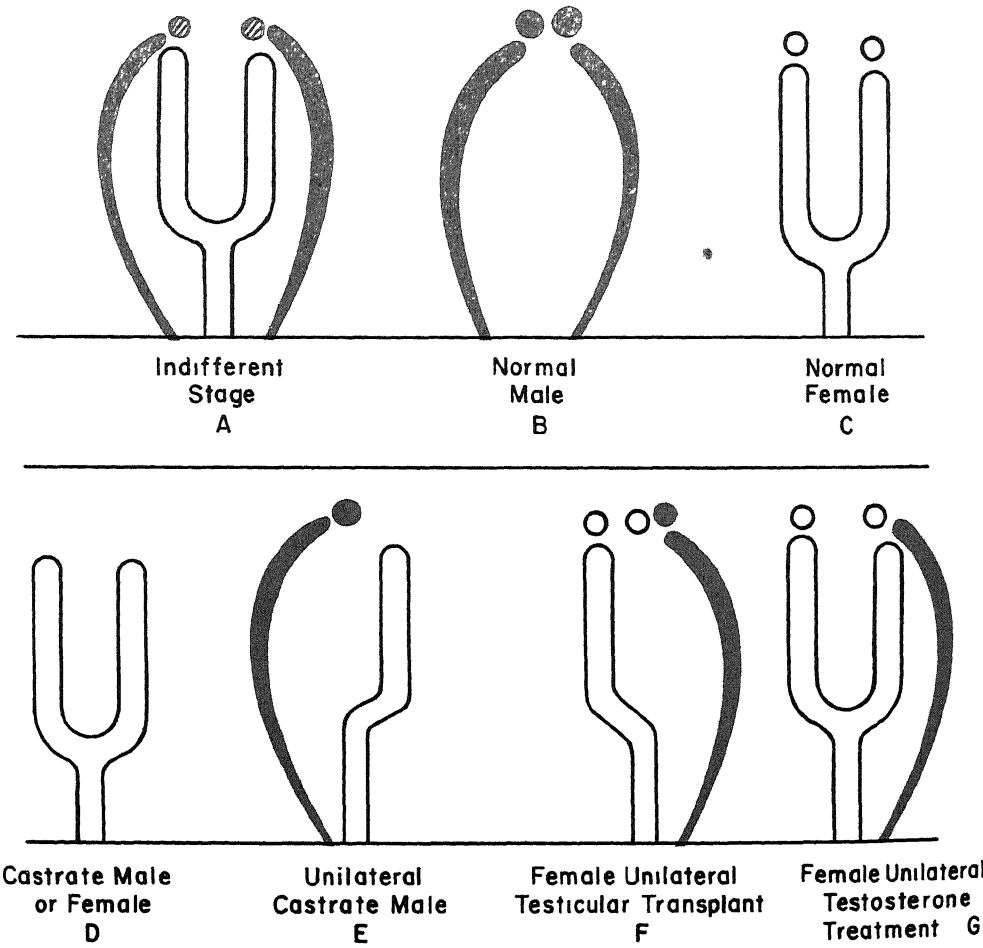


Fig 7 75 Schematic summary of Wolffian and Müllerian development in rabbit embryos (Jost, 1953) Ovaries and Müllerian structures are indicated by open figures and the Wolffian structures by the shaded areas

plant (Fig 7 75*F*) By contrast testosterone treatment of the female resulted in only Wolffian development (Fig 7 75*G*) From these observations the hormonal control of gonaduct differentiation can be summarized as follows

1) Testosterone stimulates Wolffian development and without the influence of this steroid this ductal system regresses

2) Mullerian duct system develops normally in the absence of a hormonal stimulus from either gonad but regression of this ductal system is due to a nonandrogenic hormonal product from the testis

Differentiation of the External Genitalia The external genitalia of both sexes are identical up until the end of the 8th fetal week and have the potential to differentiate into the organs of either sex At this stage the major components of the external genitalia are the genital tubercle, urethral folds, urogenital slit, and labioscrotal swelling (Fig 7 76) Under the influence of testosterone from the fetal testes (or from exogenous administration) male differentiation of the external genitalia occurs The genital tubercle forms the penis, the urogenital folds, the penile urethra, and the labioscrotal swellings, the scrotum In the absence of androgens female differentiation of the external genitalia proceeds with the development of a clitoris, labia minoria, and labia majora The structures which develop from the indifferent genital primordia in the presence and in the absence of androgens are summarized in Table 7 9 It is significant that the androgen dependent differentiation of the external genitalia can only occur during fetal life Even though the female structures are not hormone dependent, once the vagina and labia have developed they cannot be differentiated into analogous male structures even if exposed to very large quantities of androgens

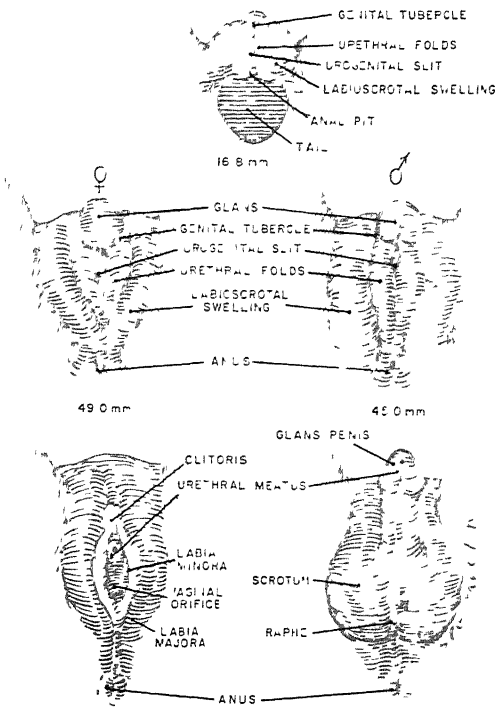


Fig 7 76 Differentiation of male and female external genitalia from indifferent primordia Male differentiation occurs only in the presence of androgenic stimulation during the first 12 weeks of fetal life (From Van Wyk and Grumbach 1968)

Hypothalamic Differentiation Even though the central nervous system is anatomically remote from the external genitalia, the differentiation of both of these structures is similar since both require fetal androgen exposure if a masculine pattern of development is to occur In the adult female, pituitary gonadotropins are released in a cyclic fashion In the mature male, gonadotropins are secreted in relatively constant rates even when an ovary is transplanted into an orchiectomized host

Table 7 9
Differentiation of the external genitalia from indifferent primordia

Primordial Structure	Structures Which Develop *with* Androgen Exposure	Structures Which Develop *without* Androgen Exposure
Genital tubercle	Penis	Clitoris
Urethral folds	Corpus spongiosum (penile urethra)	Labia minora
Labioscrotal swellings	Scrotum	Labia majora
Urogenital sinus	Prostatic utricle	Lower ⅓ of vagina
	Bulbourethral glands	Skene's glands
		Bartholin's glands

Androgen treatment of female rat fetuses either in utero or in the early neonatal period results in a noncyclic pattern of gonadotropin secretion after puberty even in the presence of a functioning ovary. From these observations it is apparent that in the absence of fetal androgens, the hypothalamus develops a cyclic center for gonadotropin release, whereas fetal androgen exposure produces a noncyclic center (Barraclough, 1966). Another way that the development of the hypothalamus is similar to that of external genitalia is that only a brief androgen exposure is required to effect an irreversible change.

Psychological Sex or Gender Role. Gender is selected at birth in accord with the apparent genital sex. This becomes the legal sex and also sets the behavioral patterns which stimulate psychosexual identification of the individual with others in society. Prior to the age of 2 years a child can identify either as a male or a female. Psychosexual differentiation or gender role is established by sex orientation through dress, hair style, toilet habits, and by constant comparison with other children and adults. Through this exposure a child is continually reminded of his gender and comes to identify with one of the sexes. From psychological studies of girls who were mistakenly raised as boys, and vice versa, it is now clear that gender role is established early in life and maintained even in the face of abnormal external genitalia and subsequent development of secondary sexual characteristics which are opposite to those of the identified sex (Hampson and Hampson, 1961). There is little evidence to suggest that gender role is an inherent characteristic of the fetus or that it is established by a direct hormone action on the central nervous system similar to that responsible for differentiation of the gonadotropin release centers.

Secondary Sexual Characteristics. The secondary sexual characteristics are the last sexual features to develop in both sexes as they occur as a result of gonadal hormonal secretion at the time of puberty. In the female, estrogens produce breast development and deposition of subcutaneous adipose tissue. By contrast, in the male testicular testosterone secretion is responsible for increase in muscle mass, deepening of the voice, male hair pattern, and growth of internal and external genitalia.

9 / HORMONAL CONTROL OF GONADAL FUNCTION

The Testis

The testis, like the ovary, has endocrine as well as reproductive functions The two major activities of the testis are segregated anatomically with androgen biosynthesis occurring in the Leydig cells and spermatogenesis in the seminiferous tubules The anterior pituitary participates in the control of both of these functions through its secretion of gonadotropins (LH and FSH) The essential features of testicular function and its control by the anterior pituitary and hypothalamus are shown schematically in Figure 7 77 Although testosterone secretion and spermatogenesis are intimately related, for convenience they are discussed independently as a hypothalamo-pituitary-Leydig cell axis and a hypothalamo-pituitary tubular axis

HYPOTHALAMO-PITUITARY-LEYDIG CELL AXIS

Luteinizing Hormone (LH) Luteinizing hormone derives its name from its function in the female where it facilitates ovulation and converts the ovulated follicle into a functional corpus luteum When this hormone was discovered in the male it was called interstitial cell-stimulating hormone (ICSH) because of its effect on testicular Leydig cells It is now recognized that these two hormones are the same, and so the term "LH" is used in both sexes As noted in Chapter 2, LH is secreted from periodic acid Schiff (PAS) positive basophils in the anterior pituitary in response to increased secretion of LH-releasing factor (LRF) from the central nervous system LH is a glycoprotein with an approximate molecular weight of 30,000 This molecule contains mannose, hexosamine, and sialic acid Both the biologic activity and the structure of LH are similar to those of human chorionic gonadotropin (HCG) which is a product of the trophoblasts of the human placenta

Leydig Cells. The interstitial cells of Leydig are located in the connective tissue stroma of the testis between the seminiferous tubules These cells differentiate in the testis of the 7-week fetus Since Leydig cells regress soon after birth it has been assumed that their initial development resulted from gonadotropin stimulation HCG is probably the hormone responsible for fetal differentiation since functional Leydig cells occur in ancephalic infants born without a pituitary (Villee, 1969) During the 1st month of neonatal life, Leydig cells revert to undifferentiated mesenchyme cells but appear once again at the time of puberty as plasma LH levels increase The morphologic differentiation of Leydig cells in the fetus, in the puberal boy, or following stimulation with either exogenous LH or HCG correlates with androgen secretion (Lipsett et al , 1966) Distinguishing morphologic characteristics of the functional Leydig cell in the human testis are abundant, smooth, endoplasmic reticulum which contains many of the enzymes necessary for steroid synthesis, lipochrome pigment, and crystalloids of Reinecke

Testicular Androgens. An androgen is a substance which stimulates growth of the prostate and seminal vesicles It is important to recognize that this is a biologic and not a chemical definition Most potent androgens are steroids and several of these are shown in Figure 7 78 Extensive studies have demonstrated that testosterone is the most important androgen both in testicular tissue and in the spermatic venous plasma Androstenedione, a testosterone precursor (Fig 7 79), is also secreted by the testis but it is a weak androgen Its only importance in men is as a plasma precursor for estrogens Other testosterone precur-

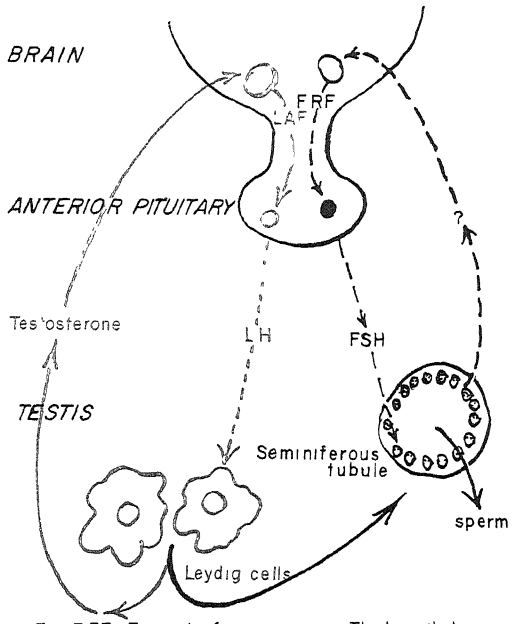


Fig 7 77 Testicular function in man The hypothalamo-pituitary-Leydig cell axis is shown on the left with pituitary LH stimulating testosterone secretion The hypothalamo-pituitary tubular axis is shown on the right with FSH and testosterone stimulating spermatogenesis A factor from the seminiferous tubules (?) is believed to regulate FSH secretion

sors such as 17-OH-progesterone (Fig 7 79) are secreted by the testis but their biologic function, if any, is not known Mean blood levels urinary levels, and production rates of testosterone and androstenedione in man are shown in Table 7 10

The Leydig cells are the primary sites of testosterone synthesis and the testes secrete 95% of the testosterone in the blood of men (Lipsett et al, 1968) In the Leydig cells, cholesterol is synthesized from acetate as in other tissues of the body In addition, mitochondrial enzymes of these cells, like those of the adrenal, ovary, and placenta, have the unique ability to cleave the cholesterol side chain to produce pregnenolone This C-21 steroid is subsequently converted to testosterone by several microsomal enzymes Some of the important steps of testicular androgen biosynthesis are shown in Figure 7 79 and are compared to steroid synthesis in the ovary

Control of Testosterone Secretion. The rate of testosterone synthesis and secretion by the Leydig cells is dependent upon LH levels in the blood (Fig 7 77) Increased LH levels produce Leydig cell hyperplasia and increased testosterone secretion, whereas lowered LH

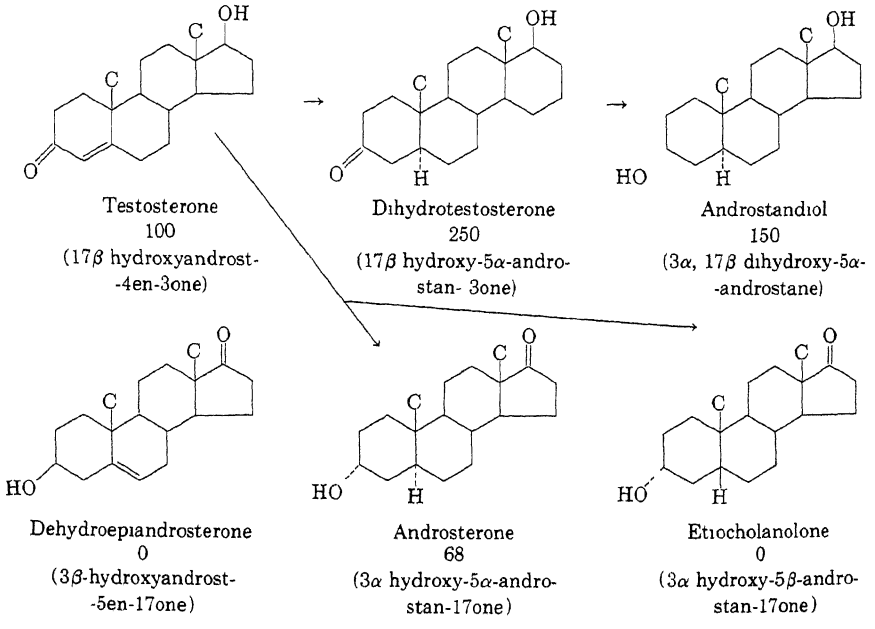


Fig. 7 78 Androgens and ketosteroids On the top line are the most important naturally occurring androgens The relative biologic potency of each is shown with testosterone having a potency of 100 Testosterone is metabolized in tracellularly to dihydrotestosterone and androstanediol On the bottom line are the relatively nonandrogenic 17 keto steroids which are secreted by the adrenal Testosterone is also metabolized to ketosteroids

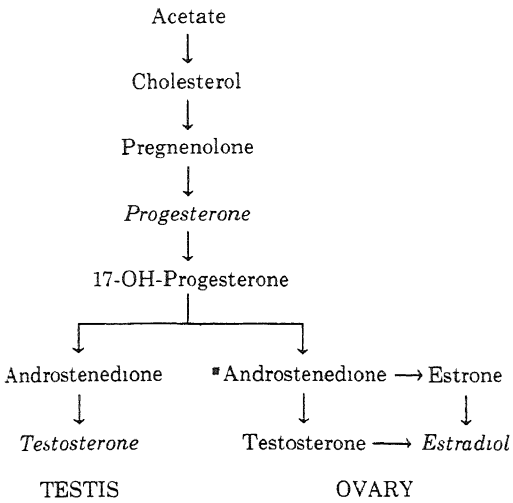


Fig 7 79 Steroid biosynthesis in the testis and the ovary Testosterone is the major secretory product of the testis The progesterone and estradiol are the major secretory products of the ovary

which follows hypophysectomy is associated with virtual inactivity of testicular Leydig cells and reduced testosterone secretion Similarly, LH secretion by the anterior pituitary is reciprocally controlled through the central nervous system by blood levels of testosterone A fall in testosterone concentration results in increased release of LRF into the hypophyseal portal system and subsequent discharge of LH into the general circulation In response to the rise in blood LH, the Leydig cells secrete testosterone which, in turn, inhibits LRF and, as a consequence, LH release (Fig 7 77) In this manner the hypothalamus, the pituitary, and the Leydig cells form a functional axis which serves to maintain testosterone blood levels relatively constant from day-to-day and from week-to-week (Lipsett et al , 1968) There is, however, a slight diurnal variation in blood testosterone with the levels at 8 00 AM being 20% to 25% higher than at 8 00 PM Resko and Eik-Nes, 1966)

Testosterone Transport and Metabolism. Steroids are organic compounds which are relatively insoluble in aqueous solutions and as a result they are transported in blood bound to plasma proteins Blood testosterone is bound with high affinity to testosterone-estradiol-binding globulin (TeBG) (Mercier-Bodard et al , 1970) This is a β globulin which is distinct from cortisol-binding globulin (Fig 7 80), the major transport protein for cortisol, corticosterone, progesterone, and other C-21 steroids. In blood 75 to 85% of testosterone is bound to TeBG whereas 15 to 25% is loosely bound to albumin and 1 to 5% is unassociated with any plasma protein (Bardin and Mahoudeau, 1970) The studies of Vermeulen et al (1969) demonstrate that in man the rate of testosterone metabolism correlates directly with the quantity of free plus albumin-bound testosterone It follows from these observations that the rate of androgen metabolism in donkey and rat are relatively high since these species lack TeBG Furthermore, these studies suggest that TeBG-bound testosterone is not readily available for in vivo metabolism It is possible that the testosterone bound to TeBG is biologically inactive since target tissue metabolism of this androgen is required for the expression of its biologic activity

Metabolism of a steroid may be defined as the irreversible removal of that steroid from the blood When testosterone enters the cell it may be metabolized either *1)* to a more active androgen which exerts its action on that cell or on a tissue in a remote part of the body, or *2)* to a less active androgen suitable for excretion either in the bile or urine (Fig 7 78) It is conceivable that both types of metabolism could occur in a sequential manner in the same cell In any event, irreversible removal from blood

Table 7 10
Testosterone androstenedione and ketosteroids in men prepuberal boys and women

	Men	Boys	Women
Testosterone[a]			
Plasma ng /100 ml	700	40	40
Urinary excretion μg /day	70	4	4
Production rate mg /day	7 0	0 3	0 3
Androstenedione[b]			
Plasma ng /100 ml	120	90	170
Production rate mg /day	2 4		3 4
Ketosteroids[c]			
Urinary excretion mg /day	12–18	2–4	8–15

[a] Testosterone, the most important blood androgen

[b] Androstenedione, a relatively nonandrogenic testosterone precursor

[c] Ketosteroids, a group of relatively nonandrogenic steroids derived from adrenal secretion and testosterone metabolism

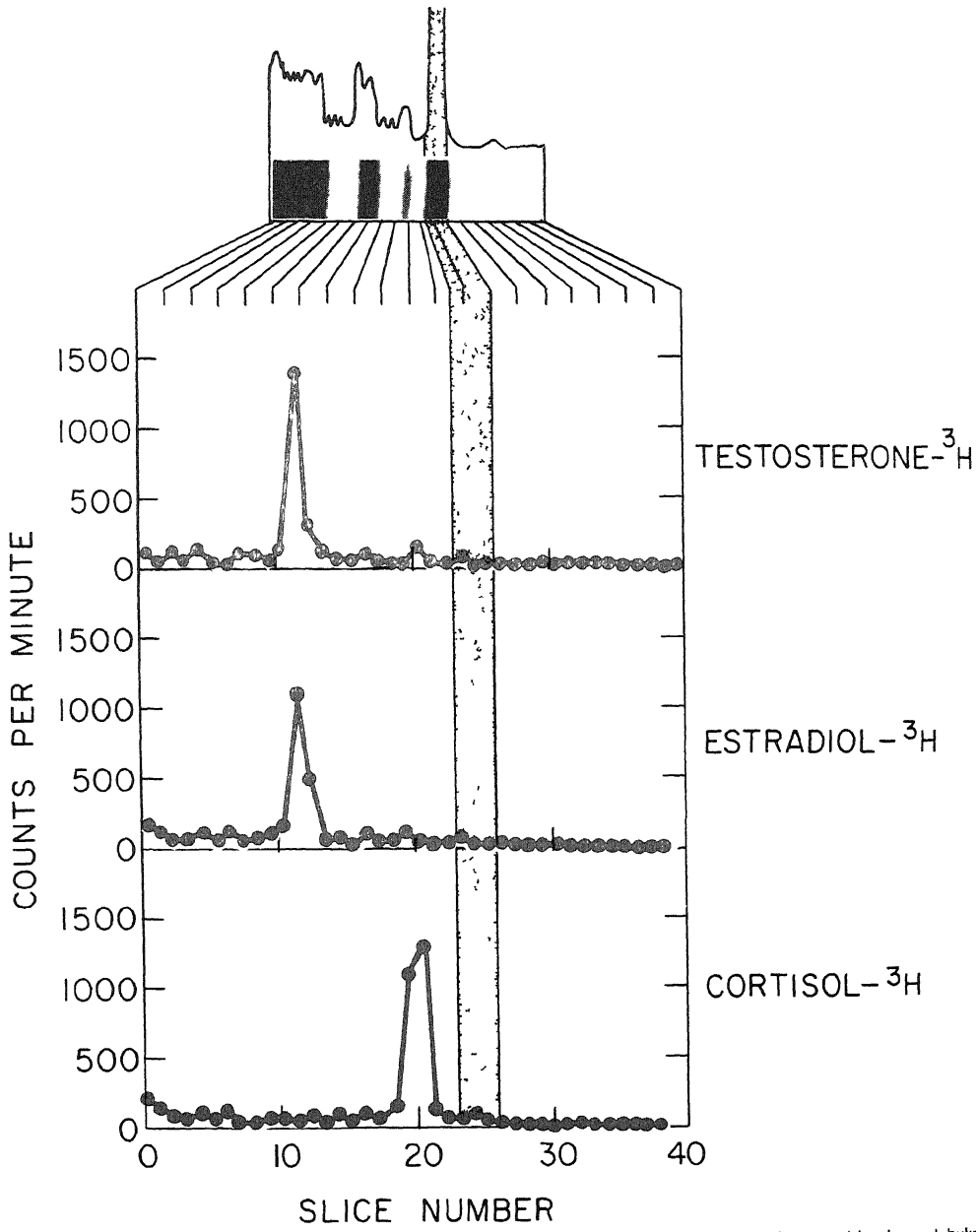


Fig 7 80 Electrophoretic demonstration of testosterone-estradiol-binding globulin and cortisol-binding globulin (bottom chart) Radioactive testosterone estradiol and cortisol were incubated with 3 samples of human plasma which was electrophoresed on polyacrylamide gels The gels were sliced and counted The peaks indicate the mobilities of the specific steroid-binding globulins The shaded bar indicates the mobility of albumin Under the conditions of the study steroid binding to albumin was not demonstrated

is accomplished in both instances by a change in steroid structure so that on re-entry into the blood is as a steroid other than testosterone In the first instance the biologic activity of testosterone is amplified and this type of metabolism occurs in tissues which grow in response to androgens such as prostate and seminal vesicles In the latter instance testosterone is metabolized to steroid sulfates and steroid glucuronides which are excreted from the

body This type of metabolism occurs primarily in the liver (Fig 7 78) (Liao and Fang, 1969)

The Biologic Actions of Androgens. Testosterone and other androgens have some biologic action on virtually every tissue in the body Their more important functions are to *1)* stimulate differentiation of Wolffian ducts, external genitalia, and the hypothalamus in the fetus, *2)* produce linear body growth, nitrogen retention, and muscular development in the adolescent, *3)* stimulate the adult maturation of the external genitalia and accessory sex organs including the penis, scrotum, prostate, and seminal vesicles, *4)* produce a low-pitched voice as a result of enlargement of the larynx and thickening of the vocal chords, *5)* stimulate beard, axillary and pubic hair growth and cause temporal hair recession and balding, and *6)* produce aggressive behavior and facilitate libido and sexual potentia In addition to these well-recognized physiologic responses, androgens stimulate or suppress selective proteins in many organs such as the liver, kidney, and the salivary glands

Mechanism of Androgen Action. Although the exact mechanism of the action for any hormone is not completely understood, considerable advances have been made in the past several years toward elucidation of macromolecular events which occur within the cell when it interacts with a hormone Androgens are thought to act directly on the nucleus of the cell to stimulate replication of nuclear ribonucleic acid (RNA) and deoxyribonucleic acid (DNA). One line of evidence supporting a nuclear site of androgen action is derived from studies of intracellular steroid localization Following an intravenous injection of radioactive testosterone, labeled steroid is retained in the nuclei of androgen sensitive tissues such as prostate and seminal vesicles long after it is cleared from the blood (Bruchovsky and Wilson, 1968) In these tissues testosterone is metabolized to dihydrotestosterone (Fig 7 78) which is bound to a nuclear receptor in the chromatin complex (Fang et al, 1969) The specificity of this receptor for dihydrotestosterone suggests that this steroid might be the intranuclear effector of testosterone action The ability of dihydrotestosterone, rather than testosterone, to enhance cytidine incorporation into nucleic acid of isolated prostatic nuclei further supports this contention (Bashirelahi et al, 1969) From these observations testosterone action on the prostate may be initiated by the following sequence of events testosterone enters the cell from the blood and is metabolized to dihydrotestosterone by 5-α-steroid reductase Dihydrotestosterone is then bound to a receptor which serves as a transport vehicle from the cytoplasm to the nucleus In the nucleus a steroid-receptor complex amplifies gene activity which results in increased nuclear RNA synthesis and eventually increased RNA directed protein synthesis (Liao and Fang, 1969)

HYPOTHALAMO-PITUITARY TUBULAR AXIS

Follicle-Stimulating Hormone (FSH). FSH is secreted from PAS-positive basophils in the anterior pituitary in response to increased secretion of FSH-releasing hormone (FSH-RH) from the hypothalamus FSH is a glycoprotein with an approximate molecular weight of 30,000 Follicle-stimulating hormone, like LH, derives its name from its function on the ovary where it stimulates follicular maturation In the male, FSH stimulates spermatogenesis as evidenced by repair of seminiferous tubules during its administration to hypophysectomized animals It should be noted that the seminiferous tubular structure can also be maintained in hypophysectomized animals by large doses of testosterone provided that its administration is begun immediately after hypophysectomy Under physiologic conditions both FSH and testosterone are required for normal spermatogenesis (Steinberger, 1971)

Seminiferous Tubules. The major mass of the testis is comprised of tightly coiled seminiferous tubules which may reach 70 cm in length The tubules are lined by spermatogonia which undergo mitotic division to give rise to the primary spermatocytes These cells mature through several stages and subsequently undergo reduction of meiotic division so that the resulting spermatids contain 23 rather than 46 chromosomes The spermatids are found near the lumen of the tubule and are transformed without further division into free spermatozoa (Fig 7 81) Approximately 74 days are required for conversion of a relatively undifferentiated spermatogonium into a highly specialized sperm cell which possesses several enzyme sys-

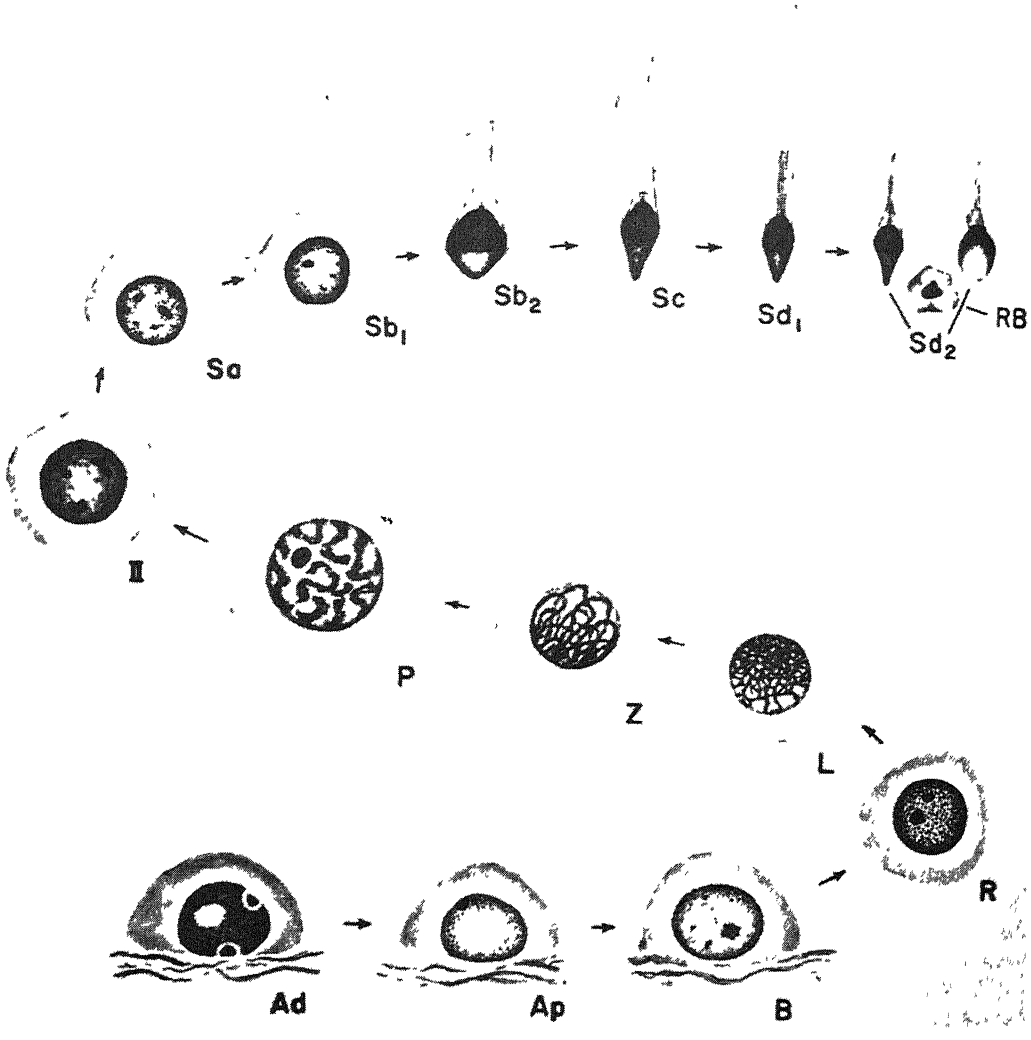


Fig 7 81 Drawings illustrating the main steps of spermatogenesis in man Spermatogenesis starts at the botton left with the dark type A spermatogonium (*Ad*) considered as the stem cell and terminates with the step d (Sd_2) spermatid or spermatozoon at the top right In sequence the other elements are *Ap* pole type A spermatogium *B* type B spermatogium *R* resting (or preleptotene) spermatocytes *L* leptotene spermatocyte *Z* zygotene spermatocyte *P* pachytene spermatocyte *II* secondary spermatocyte *Sa* Sb_1 Sb_2 *Sc* Sd_1 spermatids at various steps of spermiogenesis (From Heller and Clermont 1964)

tems that are required for penetration of the ovum (Heller and Clearmont, 1964)

The structure of the tubule is unique in that the spermatogonia are the only germinal cells which are in contact with the basement membrane The sustentacular cells of Sertoli also lie on the basement membrane of the tubules and these cells send out tentacles through the germinal epithelium so that each germinal cell is surrounded by Sertoli cell cytoplasm This anatomical arrangement has led to the conclusion that Sertoli cells nourish the germ cells as they move away from the basement membrane toward the lumen of the tubule The observation that some Sertoli cell tumors secrete estrogens has led to the conclusion that the normal Sertoli cell has a similar endocrine function There is little evidence, however, to support this conclusion in man

Control of FSH Secretion. Plasma FSH levels rise in men following orchiectomy indicating that a hormonal function from the testis

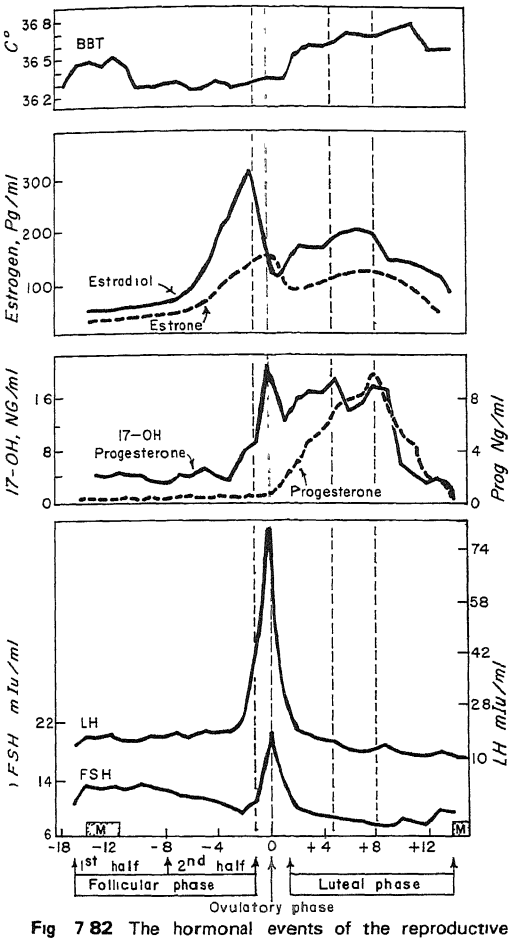


Fig 7 82 The hormonal events of the reproductive cycle are centered on the midcycle LH peak (day 0) Basal body temperatures (BBT) are shown on the top line Plasma levels of lutenizing hormone (LH) and follicle-stimulating hormone (FSH) are shown on the bottom two lines Gonadal steroid levels (estradiol estrone 17-hydroxyprogesterone and progesterone) are shown in the center Days of menstrual bleeding are indicated by M (Redrawn from Ross et al 1970 and Koreman et al 1970)

less than that of estradiol Some important steroidal and nonsteroidal estrogens are shown in Figure 7 83. The important biologic activities of estrogen include: *1*) stimulation of growth of both the myometrium and endometrium, *2*) maintenance of a thick vaginal mucosa and indirectly the acidic vaginal pH, *3*) stimulation of cervical glands to secrete copious quantities of viscous mucous, *4*) stimulation of breast growth and development, *5*) deposition of subcutaneous fat which results in a characteristic feminine habitus, *6*) sensitization of the ovaries to gonadotropins, and *7*) retardation of linear body growth in association with facilitation of epiphyseal closure

Estradiol and estrone are synthesized in the ovary from steroids with 19 carbon atoms The biosynthetic pathway up to the point of androstenedione and testosterone is essentially the same as that in the testis (Fig 7 79) In the final series of reactions the steroid A ring is aromatized with the loss of one methyl group to form the C-18 steroid estrogens (Fig 7 83) The day-to-day fluctuations of plasma estradiol and estrone during the reproductive cycle are shown in Figure 7 82

Progestins. Progestin was the name originally given to the crude extract of the corpora lutea which could prepare and maintain a secretory endometrium during the latter half of the reproductive cycle and during pregnancy Progesterone is the most active principle in these extracts Hormones which have progestational activity are referred to as progestins or gestagens (Fig 7 84)

The biologic actions of progesterone and other gestagens include *1*) antagonism of the growth promoting effect of estrogen on the endometrium and conversion of this rapidly proliferating organ into a secretory structure capable of maintaining an implanted blastocyst, *2*) conversion of the cervical mucous from a very viscous to a nonviscous fluid, *3*) stimulation of mammary gland growth and development, and *4*) inhibition of uterine motility

During the reproductive cycle and in the first trimester of pregnancy, progesterone is secreted almost exclusively by the corpus luteum During the latter two-thirds of pregnancy the placenta assumes this function The adrenal also secretes a small quantity of progesterone In all of these organs, progesterone is synthesized from cholesterol and pregnenolone

HORMONAL CHANGES DURING THE REPRODUCTIVE CYCLE

The changes in the plasma levels of estrogens, progestins, and gonadotropins during a reproductive cycle are shown in Figure 7 82 For descriptive purposes the reproductive cycle can be divided into a follicular phase (which can be subdivided into a first and second half), an ovulatory phase, and a luteal phase (Ross et al, 1970, VandeWiele, et al, 1970)

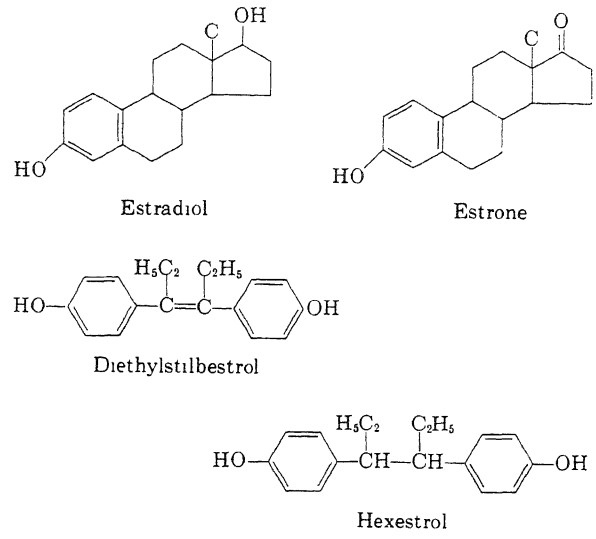


Fig 7 83 Important naturally occurring (top line) and synthetic (bottom lines) estrogens

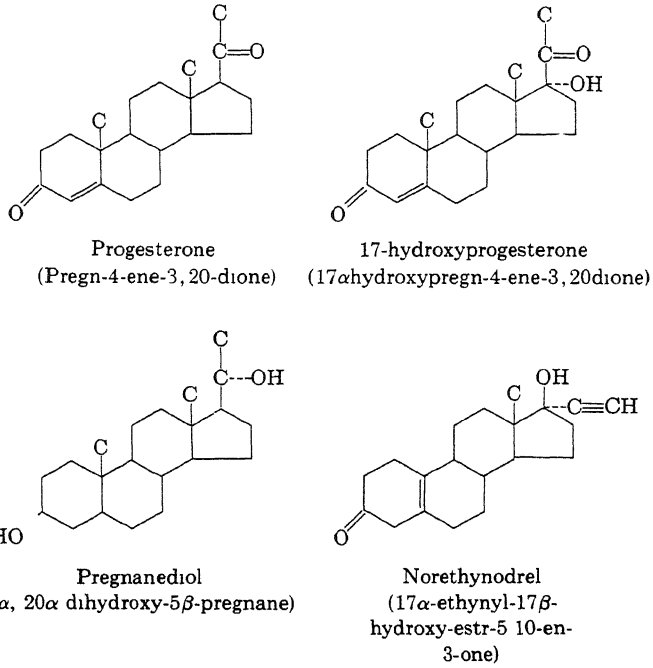


Fig 7 84 Progesterone and two of its metabolites Norethynodrel is a synthetic gestagen in clinical use

Follicular Phase, First Half. This period begins in the late luteal phase of the preceding cycle with a rise in the plasma levels of FSH and a concomitant initiation of follicular growth. LH levels also rise during this phase but the increase starts several days later than the increase in FSH. It is important to note that during the first half of the follicular phase there is no change in the blood levels of estrogens or progestins (Fig 7 82)

Follicular Phase, Second Half. This portion of the follicular phase begins approximately 7 to 8 days before the preovulatory LH surge and is characterized by an increase in

plasma estrogen levels (estradiol and estrone) Estradiol increases slowly at first and then rapidly reaches a maximum on the day before the LH peak The initial rise in estrone is not as great, but estrogen levels reach a peak concomitant with the peak of LH While plasma estrogen levels are increasing, plasma FSH decreases and plasma LH slowly and steadily increases Several days before ovulation there is a rise of 17-hydroxyprogesterone which reaches a maximum on the day of the LH surge It is significant that plasma progesterone levels do not increase during this period

Ovulatory Phase During this period there is a rapid rise in plasma LH levels which leads to the final maturation of the Graafian follicle and follicular rupture some 16 to 24 hr after the LH peak Although there is a simultaneous increase in plasma FSH levels, the role of this pituitary hormone on the ovulatory process is not understood Soon after the beginning of the LH surge and prior to ovulation, plasma estradiol levels drop precipitously and plasma progesterone increases slightly (Fig 7 82)

The preovulatory peaks of estradiol, and 17-hydroxyprogesterone correlate with rapid thecal development and follicular maturation During the final rapid growth of the follicle, the theca interna enlarges and assumes a functional appearance Following ovulation the theca interna cells persist in the corpus luteum Both estrogen and 17-hydroxyprogesterone secretion decrease rapidly in the ovulatory period but both increase again in the luteal phase as the corpus luteum with both its theca interna cells and luteinized follicular cells becomes vascularized While these observations do not prove that estrogens and 17-hydroxyprogesterone are secreted by the theca interna cells, they are nonetheless consistent with that hypothesis (Strott et al , 1969)

Luteal Phase. As noted above, luteinization is initiated by LH Recent studies also have indicated that ova exert an inhibitory influence on follicle luteinization When the ovum is removed from a mature follicle, granulosa cells immediately luteinize and form a corpus luteum By contrast, when the ovum is removed from an immature follicle, the granulosa cells degenerate and the follicle becomes atretic (Nalbandov, 1970) From these studies it was concluded that luteinization occurs spontaneously following expulsion of the ovum from the mature follicle since only the ovectomized follicles luteinized and the neighboring follicles remained unchanged These observations are consistent with those of Channing (1969) who observed that follicular cells placed in tissue culture luteinized spontaneously in the absence of all hormonal stimuli However, the lifespan and function of follicular cells which are luteinized spontaneously either in vivo or in vitro are reduced LH is required for normal corpus luteum survival and function Furthermore, LH administered in high doses can luteinize follicular cells even without ovulation and can produce a luteinized follicle with a trapped ovum (Ludwig and Horowitz, 1969) These observations indicate that either LH or ovum removal can luteinize the ovarian follicle, however, during a normal reproductive cycle both mechanisms are probably operative

The *sine qua none* of the luteal phase is the marked increase in progesterone secretion which reaches a maximum about 8 days after the midcycle LH peak 17-hydroxyprogesterone has a double peak during the luteal phase at 5 and 8 days, respectively, and as noted previously, estrone and estradiol plasma levels increase in parallel with those of the progestins Concomitant with the increase of gonadal steroids, plasma LH and FSH decline throughout most of the luteal phase (Fig 7 82)

The factors which control the lifespan of the corpus luteum in the normal reproductive cycle are ill-defined In many species, termination of luteal function is under control of a luteolytic factor produced by the uterus In the human and monkey there is no evidence for the existence of such a factor (Rowson, 1970)

Control of FSH and LH Secretion. Both FSH and LH are subject to negative feedback control as evidenced from the rise in both of these hormones following castration and from the decrease in their levels following estrogen administration Examination of the hormonal events of the reproductive cycle indicates that control mechanisms in addition to "negative feedback inhibition" are operative This is particularly true in the second half of the follicular phase when FSH decreases and LH increases concomitant with a rapid rise of blood estrogens. This paradox may be explained if one assumes that there are hypothalamic centers for positive and negative feedback control of gonadotropins in man as there are in the rodent (McCann and Porter, 1969) In addition,

to the negative feedback upon gonadotropin secretion, estrogens also exert a positive feedback on LH secretion but not upon FSH The factors which determine whether LH will increase or decrease in response to estrogen administration are not completely understood at the present time Suffice it to say that from the available data it would appear that both a positive and a negative feedback control of LH are operative at different periods during the reproductive cycle

In light of these observations the difference between the behavior of FSH and LH during the late follicular and luteal phases may be rationalized In the second half of the follicular period as estrogens rise, FSH decreases due to the negative feedback effect By contrast, the positive feedback control of estrogen is operative upon LH secretion during this period as the level of this hormone increases Both FSH and LH decline during the luteal phase under the negative feedback inhibition of estrogen and possibly progesterone

In addition to the feedback control of gonadotropins by gonadal steroids, there are also mechanisms by which gonadotropins control their own secretion This latter type of gonadotropin control has been termed the short-loop feedback while the control mechanisms involving gonadal steroids is called the long-loop feedback The following observations in rodents suggest that LH may exert a negative feedback on its own secretion via a "short loop" *1*) implantation of LH in the median eminence of rats results in decreased pituitary LH stores and decreased plasma LH levels, *2*) LRF increases following hypophysectomy (David et al , 1966) At present there is not conclusive evidence for a negative short loop feedback of FSH control (Ojeda and Ramirez, 1969) Although short-loop feedback is assumed to be operative in primates, there are no data to confirm or negate the existence of this type of feedback in man

A midcycle LH surge represents still another facet of the complex machanisms which are operative in the control of gonadotropin secretion in women The available studies support the hypothesis that the rise in estrogens during the later follicular phase triggers the preovulatory LH surge Evidence in favor of this theory has been provided by several investigators who demonstrated that estrogen administration can produce repeated bursts of LH, and in some instances, ovulation Further evidence for estrogen-stimulated LH release is obtained from experiments using antibodies to estradiol in which LH secretion was blocked when the biologic activity of estradiol was neutralized (VandeWiele et al , 1970) Under certain conditions progesterone can stimulate LH release (Swerdloff and Odell, 1969), but a careful analysis of the sequence of hormonal events during the reproductive cycle indicates that progesterone does not rise prior to LH (Fig 7 82) (Ross et al , 1970) Recent studies with progesterone antibodies are also consistent with these observations (Ferin et al , 1969)

Transport and Metabolism of Estradiol and Progesterone. The transport of estradiol and progesterone in blood is similar to that of testosterone A large fraction of estradiol in blood is bound with high affinity to TeBG (Fig 7 80) with smaller fractions associated with albumin and red cells Progesterone and cortisol bind with approximately equal affinities to cortisol-binding globulin (CBG) Since cortisol always circulates in at least a 10-fold greater concentration in plasma than progesterone, relatively more of the binding sites on CBG will be occupied by cortisol than by progesterone Therefore, even though progesterone can bind to a specific plasma transport protein, the majority of progesterone is transported by albumin since the specific binding sites are occupied by another steroid

and a variety of 2-methoxy compounds Estradiol and its metabolic products are conjugated with sulfate and/or glucuronic acid for excretion The keto groups of progesterone are reduced to form pregnanediol (Fig 7 84) which is then excreted as a conjugate Unlike testosterone, estradiol and progesterone are not metabolized to more active steroid products as both of these hormones are biologically active on target tissues in the form in which they are secreted into the blood

Origin of Testosterone in Women The origin of testosterone in women is illustrated in Figure 7 85 and is analagous to the production of estrogens in man Testosterone is secreted into the blood stream from both the ovary and the adrenal cortex Both of these endocrine glands, in addition, produce relatively nonandrogenic steroid precursors or prehormones such as androstenedione and dehydroepiandrosterone Steroid prehormones are then converted to testosterone in liver and pe-

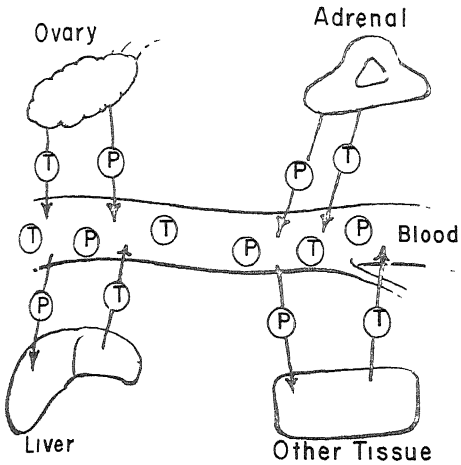


Fig 7 85 The origin of blood testosterone in women Testosterone (*T*) is secreted into the blood by the ovary and adrenal These endocrine glands also secrete precursor steroids or prehormones (*P*) into the blood which are transported to the liver and other tissues where they are converted to testosterone which re-enters the blood

ripheral tissue In normal women androstenedione and dehydroepiandrosterone serve as precursors for 50 and 15%, respectively, of plasma testosterone Since the origin of blood testosterone in normal women is diverse, one can anticipate similar but variable diversity of testosterone production in women with virilizing disorders (Bardin and Mahoudeau, 1970)

Mechanism of Estrogen and Progesterone Action Estradiol and progesterone, like other steroids, are thought to exert their action directly on the nucleus of the cell Using sucrose gradients, Jensen and his colleagues (1968) have demonstrated the sequence of events leading to estradiol interaction with the nucleus of the uterine cell Upon entry into the cell, estradiol is bound to an 8S receptor protein in the cytosol The estrogen is transferred to the nucleus, and in the process the cytoplasmic receptor is altered In the nucleus, estradiol is bound to a 5S receptor and this steroid-receptor complex is believed to activate DNA and RNA synthesis

Progesterone receptors have also been demonstrated in the uterus of a variety of animals These receptors are unique in that their activity is low in premature and castrate animals and is increased by estrogen treatment If a cytoplasmic receptor for progesterone is a prerequisite for progesterone action on the cell, then its maintenance by estrogen can explain the synergism of these two hormones on the uterus

PREGNANCY

Development of the Placenta In the human, fertilization occurs in the fallopian tube There the early zygote divides repeatedly as it makes its way into the uterine cavity 3 or 4 days after fertilization By the time the zygote reaches the uterine cavity it is a solid ball of cells called a morula The morula accumulates fluid between the cells which eventually causes separation of an outer layer of trophoblastic cells and an inner layer of embryonic cells that are attached to the trophoblastic layer in a restricted area The embryo, with its protective layer of trophoblasts, is termed the blastocyst, and by the 7th day after fertilization it is ready for implantation By the 9th day after fertilization, it is possible to detect human chorionic gonadotropin (HCG) in maternal urine This protein is synthesized by the trophoblast and has a hormonal role in the maintenance of the corpus luteum of pregnancy This hormone is made solely by the placenta so that hypophysectomized, pregnant women have normal excretions of HCG Thereafter the placenta continues to mature and serves both endocrine and nutritive functions It is capable of steroid synthesis by 6 weeks and its definitive structure is attained by 12 weeks of pregnancy (Villee, 1969)

Hormones of Pregnancy. During pregnancy the placenta is responsible for the production of a number of steroid and protein hormones which exert a variety of metabolic effects both on the fetus and mother

Progesterone is required throughout pregnancy to maintain a functional rather than a proliferative endometrium During the first few days of pregnancy (from fertilization until just after implantation), the progesterone required for endometrial maintenance is derived from the corpus luteum which is under the control of pituitary LH This control will last approximately 10 to 13 days, and, unless placental HCG is produced in sufficient quantities to maintain the corpus luteum beyond this period, the pregnancy will terminate several days after implantation as a consequence of luteolysis, a fall in progesterone concentration, and menstruation In normal pregnancy, however, corpus luteum maintenance is transferred from pituitary to placental control and progesterone

secretion continues Both progesterone and 17-hydroxyprogesterone increase until peak concentrations are attained at 3 to 4 weeks of gestation (Fig 7 86) Subsequently both steroids decline at the same rate, the plasma progesterone level reaching a nadir at 6 to 8 weeks after which it increases again as a result of placental secretion By contrast, 17-hydroxyprogesterone levels continue to decline Since the placenta does not synthesize significant quantities of 17-hydroxyprogesterone, this steroid is a good index of corpus luteum function during pregnancy These data suggest that the corpus luteum of pregnancy functions maximally during the first 4 weeks following ovulation and that its lifespan may be only about 10 weeks (Yoshimi et al , 1969)

Although the placenta possesses all the enzymes required to form cholesterol from acetate, it appears that maternal plasma cholesterol is the principal precursor of progesterone produced in the placenta Since the placenta cannot 17-hydroxylate progesterone, androgens and estrogens cannot be synthesized directly in the placenta as in the ovary and testis By contrast, these steroids are synthesized in the placenta from steroid precursors produced in the fetus Participation of the fetus and the placenta in the synthesis of hormones has given rise to the concept of the fetoplacental unit which is discussed below

Human Chorionic Gonadotropin is a glycoprotein hormone produced by the trophoblasts of the placenta HCG is chemically and biologically similar to pituitary LH and is believed to maintain the early corpus luteum of pregnancy The level of this hormone rises in maternal blood throughout the early weeks of pregnancy and peaks at 9 weeks when the number of cytotrophoblasts is maximal The fact that plasma progesterone levels decline, indicating involution of the corpus luteum of

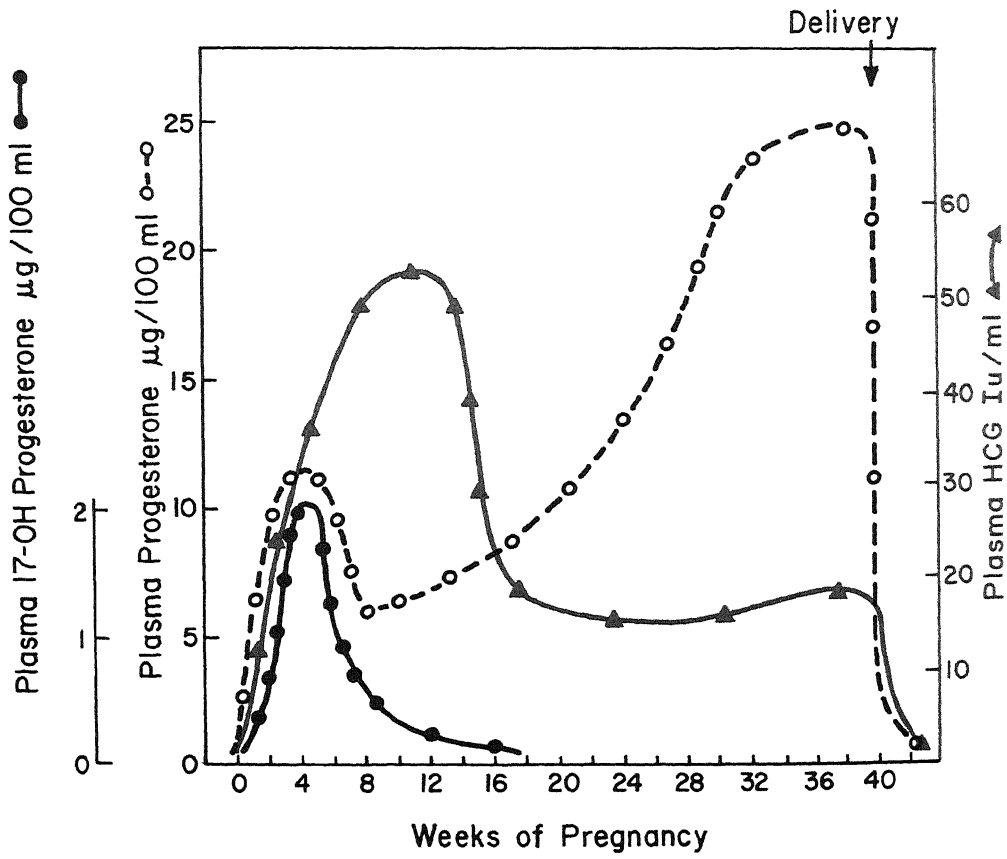


Fig 7.86 Maternal plasma levels of 17-hydroxyprogesterone progesterone and human chorionic gonadotropin (HCG) during pregnancy

pregnancy, while HCG levels are rising indicates that factors other than blood HCG levels are responsible for the duration of corpus luteum function After 9 weeks, plasma HCG levels also decline (Fig 7 86)

Human Placental Thyrotropin is a glycoprotein produced in the human placenta which is chemically similar, but immunologically distinct, from pituitary thyrotropin The existence of this hormone was first demonstrated in women with trophoblastic tumors of placental origin who developed thyrotoxicosis The fact that anencephalic infants have a marked reduction in thyroid size suggests that this hormone does not significantly influence thyroid function of the fetus in utero (Hertz, 1964)

Human Placental Lactogen (HPL) is another placental hormone produced by the trophoblasts as early as 3 weeks after ovulation Very little of this hormone can be detected in a fetus but very large quantities are present in the maternal circulation HPL can be detected in serum of pregnant women as early as 6 weeks of gestation and increases steadily during the first 6 months Blood levels show little day-to-day fluctuation and are not influenced by changes in blood glucose concentrations HPL disappears from the blood promptly after the delivery of the placenta

HPL has prolactin activity on the pigeon crop gland and has diabetogenic properties on mammals Although the exact role of this hormone in human pregnancy is not understood, a number of metabolic changes during gestation may be due to HPL These include hyperinsulinema, accelerated lipolysis, elevation of plasma free fatty acids, inhibition of glucose uptake, and acceleration of gluconeogenesis (Grumbach et al , 1968)

Fetoplacental Units. After conception there is a dynamic nutritional interplay between the fetus and the placenta which assures their mutual growth and development In addition, many studies indicate that the fetus and placenta function as a unit for synthesis and metabolism of steroid hormones This conclusion is supported by the fact that both the fetus and placenta have incomplete steroidogenic systems so that each complements the other in the process of estrogen biosynthesis Large quantities of dehydroepiandrosterone sulfate and 16-hydroxydehydroepiandrosterone sulfate are synthesized by the fetus Both steroids are transported to the placenta where the sulfate is removed (Fig 7 87) Dehydroepiandrosterone is then converted to androstenedione and testosterone which are aromatized to estrone and estradiol, respectively 16-α-hydroxy dehydroepiandrosterone is converted to 16-α-hydroxy testosterone which is then aromatized to estriol Estrogens synthesized by the fetoplacental unit enter the maternal circulation and comprise a major fraction of the estrogens of pregnancy From an analysis of Figure 7 87 it is easy to understand why fetal death is associated with a decline in maternal estrogen levels

Breast Development The secretory tissue of the breasts or mammary glands consists of elongated slender ducts whose walls have numerous pouches—the alveoli The walls of the alveoli are composed of a basement membrane, a layer of myoepithelial cells, and an inner layer of columnar epithelium The epithelial cells are the secreting elements of the breasts

There are three stages of phases of breast development Prior to puberty, breast growth is isometric to that of body growth At puberty, both the ductal system and the supporting adipose tissue grow much faster than the body in general due to increased secretion of ovarian estrogens Women without ovaries, therefore, develop breasts of normal size and consistency only when treated with estrogens The final stage of breast development occurs during pregnancy This phase is initiated by rapid growth of ducts and alveoli which differentiate so that milk synthesis and secretion can occur While the hormone-dependent growth of breast tissue during puberty appears relatively straightforward, breast growth during pregnancy is exceedingly complex

Surprisingly little is known about the hormonal requirements in women for breast growth and differentiation during pregnancy However, the hormonal requirements for growth of mouse mammary gland and milk synthesis have been well-defined in tissue culture and indicate the need for multiple hormones Growth of ducts and alveoli to midpregnancy can be achieved by administration of estrogen and progesterone to the rodent provided the pituitary and adrenals are intact These organs provide growth hormone and glucocorticoid, respectively, which are required for mammary gland growth If the mammary gland of an animal in midpregnancy is placed

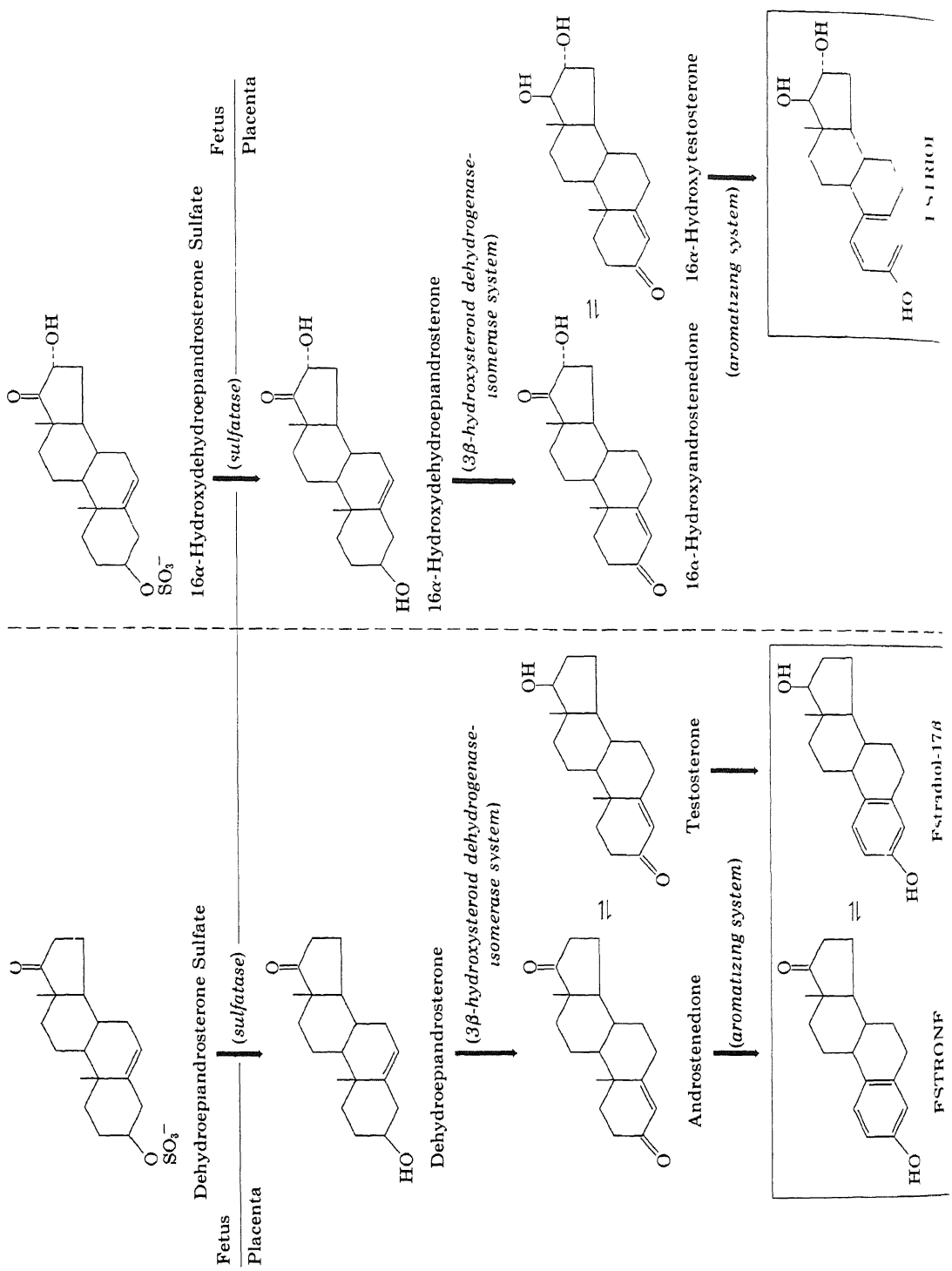


Fig 7 87 Synthesis of estrogen by the human placenta The pathways for the conversion of C19 steroid sulfates to estrogens are shown The major estrogens secreted by the placenta are boxed in (From Villee 1969)

in tissue culture, insulin, glucocorticoid, and prolactin are required to produce mammary gland differentiation and milk synthesis (Topper, 1970) While there are species differences in the hormonal requirement for mammary growth and development, gonadal adrenal, and pituitary hormones have been necessary in all animals studied to date

Although milk synthesis and secretion in women begins about the 5th month of pregnancy, the initiation of copious lactation is a consequence of delivery Milk release or "letdown" from the breast at the time of nursing is accomplished by the action of oxytocin on the myoepithelial cells of the alveoli Oxytocin is in turn produced in the hypothalamus and is released from the neurohypophysis

7 CONTROL OF CARBOHYDRATE METABOLISM IN MUSCLE

In muscle tissue, the rate of glycolysis varies with the degree of mechanical activity, with the nutritional or hormonal state of the animal, and with the type and function of the muscle For example, white skeletal muscle has few mitochondria, contracts very rapidly but can sustain a continuous contraction for only a short time This type of muscle depends on glycolysis as a primary source of adenosine triphosphate (ATP) Red skeletal muscle contracts slower but for a longer time, is rich in mitochondria, and derives most of its ATP through oxidation of fatty acid and glucose Cardiac muscle is a type of red muscle Mitochondria make up about 40% of the total volume of cardiac muscle and, under normal conditions, oxidative metabolism accounts for almost all of the ATP that is produced Glycolysis becomes an important source of ATP in red muscle only when the oxygen supply is limited, as during strenuous exercise in skeletal muscle or coronary occlusion in cardiac muscle

Under most conditions, glucose and fatty acids are the major fuels used by muscle, but lactate, pyruvate, ketone bodies and, to a minor extent, amino acids are all potential sources of energy In a well-oxygenated red muscle, all of these substrates are completely oxidized in the citric acid cycle resulting in formation of reduced nicotinamide adenine dinucleotide (NADH) which serves as the immediate energy source for ATP production by oxidative phosphorylation The rate of utilization of any one substrate depends upon the concentration of the substrate in the plasma, availability of alternative substrates, degree of mechanical activity, availability of oxygen, and plasma levels of hormones When glucose, insulin, fatty acids, and oxygen are present in normal amounts in arterial blood, oxidation of glucose accounts for only 20 to 30% of the total oxidative metabolism of the heart, while oxidation of fatty acids accounts for the remainder When insulin levels are low and fatty acids are high, as in fasting or uncontrolled diabetes, the heart utilizes fatty acids and ketone bodies almost exclusively On the other hand, when blood glucose and insulin are high, as after a carbohydrate meal, oxidation of glucose may account for a larger portion of cardiac metabolism Utilization of one substrate in preference to another indicates that regulation of overall metabolism involves a complicated network of interactions of one pathway with another which brings about an integration of all pathways, and insures that the rate of each is coordinated with the metabolic needs of the cell as a whole

Although this chapter is concerned with the regulation of glucose metabolism in muscle, many of the regulatory mechanisms that are discussed apply to other tissues as well Glucose utilization by adipose tissue and the influence of glycolysis on fatty acid release are discussed in Chapter 11

General Considerations of Metabolic Regulation

In general, complicated metabolic pathways are regulated as units A unit often represents a sequence of reactions between major branch points in a pathway For example, glucose metabolism appears to be regulated as three units (Fig 7 88) The glucose-activation unit is comprised of membrane transport and intracellular phosphorylation of glucose to glucose-6-P Glucose-6-P represents the first branch point in the pathway and may be used for glycogen synthesis, for synthesis of other metabolites via the pentose cycle, or for production of ATP by conversion to pyruvate or acetyl CoA through the main glycolytic pathway Glycogen metabolism is a cyclic process involving both synthetic and degradative enzyme systems Functionally, both enzyme systems are regulated as a glycogen homeostatic unit

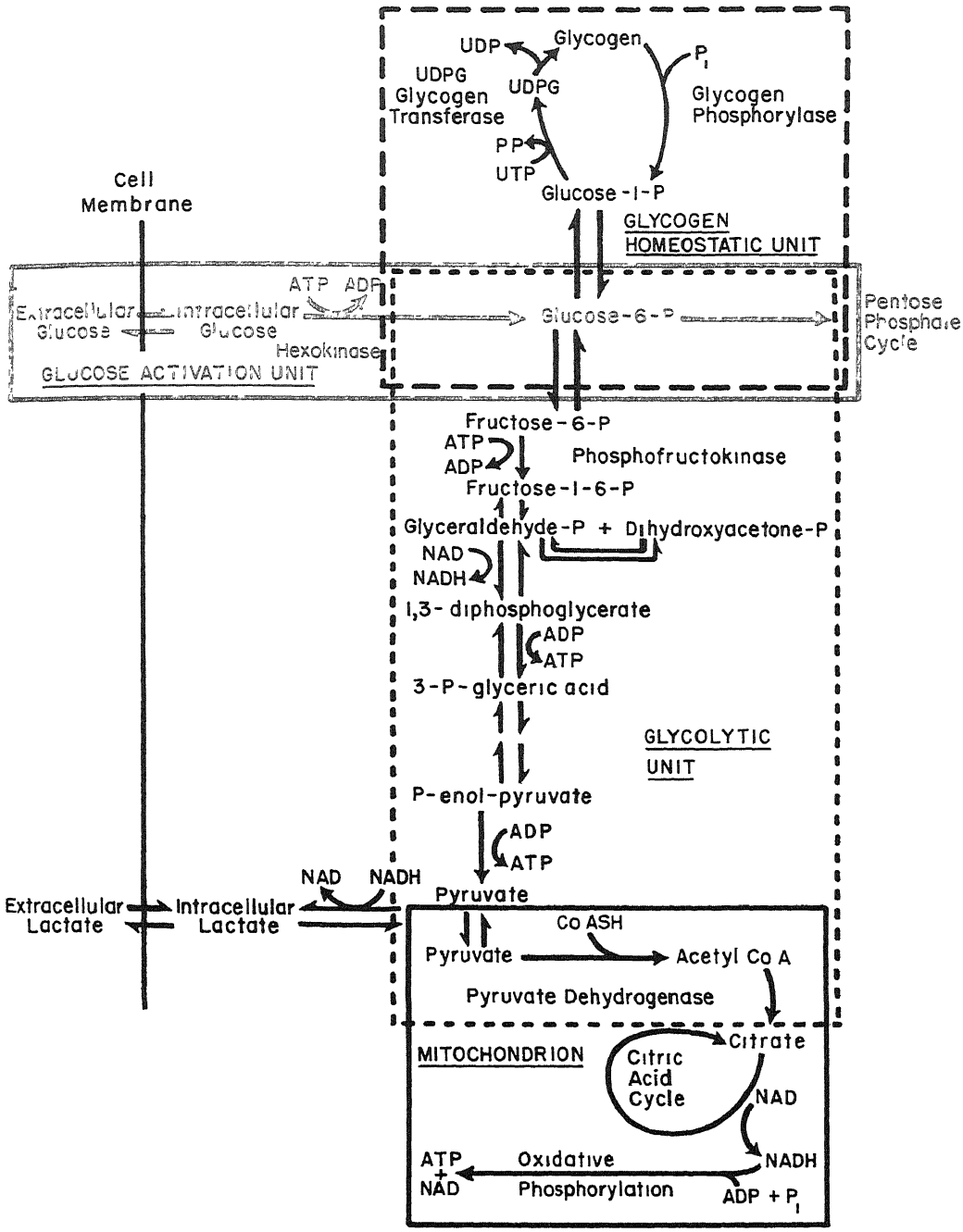


Fig 7 88 Glycolytic control units

When synthesis is stimulated, glycogenolysis is inhibited and visa versa Acetyl CoA represents the next major branch in glycolysis after glucose-6-P and the series of reactions from glucose-6-P to acetyl CoA is regulated as a glycolytic unit Acetyl CoA can be oxidized through the citric acid cycle or can be used for synthetic processes Conversion of pyruvate to lactate does not represent a major branch point in glycolysis Lactate production in anaerobic muscle provides a mechanism for oxidizing the NADH produced at the glyceraldehyde-3-P reaction and allows glycolytic production of ATP to continue in the anaerobic

cell Lactate formed by anaerobic muscle is transported to the liver where it is converted to glucose by gluconeogenesis In aerobic tissue, the lactic dehydrogenase reaction can be reversed and the pyruvate that is produced is oxidized via the citric acid cycle

"Metabolic units" are regulated by control of one or more enzymes within the sequence of reactions, often through feedback inhibition of either the first step or the first irreversible step in the sequence The inhibitor may be the immediate product of the reaction or a more distant product of the pathway For example, hexokinase is inhibited by glucose-6-P, the immediate product of the reaction, but phosphofructokinase is inhibited by ATP, the ultimate product of glycolysis and of aerobic metabolism This enzyme is also inhibited by citrate which is synthesized from acetyl CoA and is therefore a product of glycolysis Pyruvate dehydrogenase is inhibited by acetyl CoA, the immediate product of the reaction, and by ATP, the final product of pyruvate oxidation

TYPES OF ENZYME CONTROL

Flow through a series of reactions may be controlled by several factors *1*) Availability of the initial substrate can regulate the overall rate Glucose utilization by muscle can be controlled to a significant extent by the plasma concentration of glucose The plasma concentration is in turn regulated by dietary intake and hepatic and renal production of glucose In addition, the rate of membrane transport can be modified by insulin and metabolic factors at any level of plasma glucose *2*) An important factor regulating flux through a metabolic pathway is the presence of enzymes that are under allosteric control and are strategically located within the pathway Regulatory enzymes of this type control the overall flux through a pathway or unit and also control the concentrations of intermediates within the pathway Phosphofructokinase often controls the overall rate of glycolysis and regulates the tissue levels of fructose-6-P and glucose-6-P. Regulatory enzymes catalyze nonequilibrium reactions, but not all nonequilibrium reactions are regulated A nonequilibrium reaction may be identified in the tissue by measuring the concentrations of all substrates, cofactors, and products of the reaction The ratio of the algebraic product of all concentrations of reaction products to the algebraic product of all concentrations of substrates is then compared with the same ratio measured in vitro for the purified enzyme (i e , the apparent equilibrium constant) If the in vivo ratio is less than the apparent equilibrium constant measured in vitro, the enzyme is said to catalyze a nonequilibrium reaction This indicates that some factor (allosteric regulator) present in the tissue is influencing the activity of the enzyme On the other hand, if a reaction proceeds at equilibrium, the rate increases when there is more substrate or less product and is slower when there is less substrate or more product (mass action effect), and cannot control the overall rate of a pathway *3*) Flux through a metabolic pathway may be changed by variations in the absolute amount of one or more enzymes in the pathway This mechanism appears to be important in long-term control of metabolic pathways in certain tissues In muscle, however, total activities of most of the regulatory enzymes that have been studied do not change under conditions where large changes in the rate of the reactions occur These findings indicate that carbohydrate metabolism in muscle is regulated primarily by availability of substrate and control of nonequilibrium reactions

REGULATORY REACTIONS OF THE GLUCOSE-ACTIVATION UNIT

Movement of glucose across muscle cell membranes occurs by a process called carrier-mediated transport or facilitated diffusion Glucose transport is freely reversible and serves to equilibrate intracellular and extracellular concentrations of glucose Since metabolic energy is not used, this process is an example of passive transport Active transport moves substances against a concentration gradient with the expenditure of metabolic energy Glucose transport is active only in the kidney and intestine

Passive transport of glucose is illustrated schematically in Figure 7 89 In this model, free glucose combines with a mobile protein component of the cell membrane ("carrier") The carrier itself can move through the membrane and is freely accessible to glucose on either side When glucose combines with the carrier on the outside of the cell, the complex renders glucose sufficiently lipid soluble to allow the sugar-carrier complex to pass through the lipid bilayer portion of the mem-

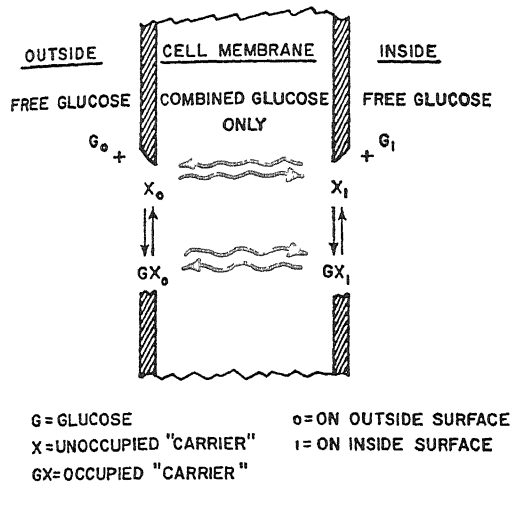


Fig 7 89 A model of passive glucose transport (From Park et al 1968)

brane The complex then dissociates and free glucose is released on the inside of the cell Intracellular glucose can combine with free carrier on the inner surface of the membrane and be transported out of the cell

Several types of experiments support this model of carrier transport *1*) Artificial lipid bilayer membranes formed from lipids extracted from cell membranes allow glucose to penetrate at a rate several hundred times slower than movement of sugar through a biological membrane Simple diffusion through the lipid bilayer, therefore, cannot account for the observed rates of glucose movement into living cells *2*) Glucose transport can be saturated as increasing concentrations of sugar are supplied This indicates that glucose combines with a component of the membrane that is present in limited amount When all of the carriers are occupied, transport is limited by the number of carriers and their mobility in the membrane *3*) Glucose transport is stereospecific as indicated by the rapid entry of D-glucose into muscle cells and by the exclusion of the L-isomer These observations also provide support for the suggestion that the carrier is a protein since proteins can selectively bind stereoisomers of small molecules *4*) Glucose transport is inhibited by sulfhydryl-blocking agents suggesting that a protein is involved in the penetration process *5*) Several sugars share the same carrier system and compete with each other for the available binding sites *6*) Countertransport, a special case of competitive inhibition, provides the strongest support for carrier mobility This phenomenon is demonstrated by allowing intracellular and extracellular concentrations of a nonmetabolized sugar, such as 3-O-methyl glucose, to equilibrate across the cell membrane and then by adding a metabolizable sugar, such as glucose, to the ouside of the cells Following addition of glucose, 3-O-methyl glucose is transported out of the cells against a concentration gradient (Fig 7 90) If both glucose and 3-O-methyl glucose are included in the buffer from zero time, entry of 3-O-methyl glucose is slower indicating that these sugars compete for the same transport system This phenomenon can be understood by reference to the carrier model (Fig 7 89) When the extracellular compartment contains only 3-O-methyl glucose, this sugar combines with the carrier and is transported into the cell where it accumulates As the concentration of 3-O-methyl glucose builds

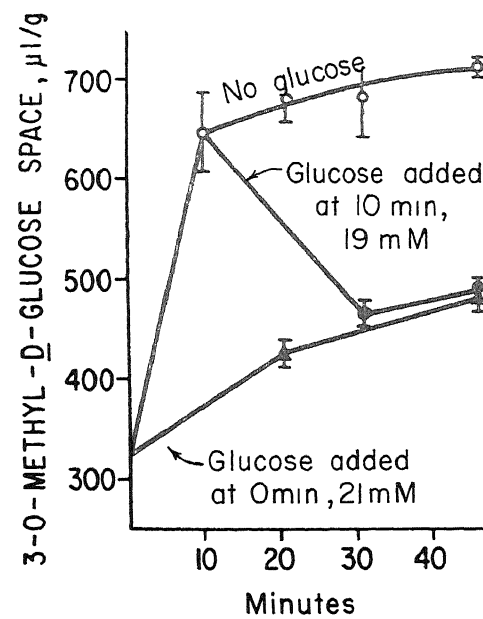


Fig 7 90 Countertransport of 3-O-methyl D-glucose in the isolated perfused rat heart on addition of D-glucose to the perfusate Hearts were perfused for the times indicated with buffer containing 3-O-methyl D-glucose-^{14}C (0 75mM) In one group of hearts perfused with 3-O-methyl glucose perfusion was switched after 10 min to buffer containing 21 mM D-glucose in addition and perfusion was continued for either 20 or 35 min (●) Additional groups were perfused with 0 75 time (▲) The vertical line through each point indicates two standard errors of the mean (From Morgan et al 1964)

up inside the cell, more sugar is available for outward transport When the concentrations are equal on both sides of the membrane, outward transport equals inward transport If glucose is now added to the extracellular compartment, it competes with 3-O-methyl glucose for binding sites on the outer surface of the membrane and reduces the inward movement of 3-O-methyl glucose As glucose is transported into the cell, however, it is phosphorylated and is not available to compete with 3-O-methyl glucose for outward transport Since more free carriers are available for binding of 3-O-methyl glucose on the inner side of the membrane than on the outer, 3-O-methyl glucose exit is faster than entrance, which results in movement of the nonmetabolized sugar against a concentration gradient Energy for this movement comes from the glucose concentration gradient

Transport is a major rate-limiting step for glucose utilization and control at this point often regulates the entire pathway In the absence of insulin, extracellular glucose can be abnormally high and yet the intracellular glucose concentration remains so low that it cannot be measured Under these conditions, transport is a nonequilibrium reaction and limits overall glucose utilization since hexokinase can phosphorylate glucose only as fast as transport supplies the sugar When transport is stimulated by insulin, free glucose may build up in the cell, which allows transport to operate near equilibrium and shifts the restraint on glucose activation to the hexokinase reaction Transport is regulated by a number of factors in addition to insulin as illustrated in Table 7 11) In these experiments, the rate of equilibration of intracellular water with 3-O-methyl glucose during 7 min of perfusion functions as an index of transport rate Transport occurs at a slow rate in aerobic hearts that are developing low levels of ventricular pressure (60 mm Hg) An even lower rate is seen when palmitate is added to the perfusate The increase in muscular work associated with greater ventricular pressure development (100 mm Hg) accelerates 3-O-methyl glucose entry, but this acceleration is blocked if palmitate is present in the perfusate Insulin accelerates sugar transport, but the hormone effect is less if palmitate is present. Perfusion of the heart under anaerobic conditions accelerates entry of sugar into the cells. Palmitate does not interfere with the anaerobic stimulation of transport which suggests that fatty acid must be oxidized to inhibit this step in glucose metabolism These studies indicate that transport is regulated by insulin, muscular work, availability of fatty substrates, and degree of oxygenation of the heart These factors interact to produce a fine control of transport rate

Table 7 11
Effects of ventricular pressure development substrates anoxia and insulin on 3-O-methyl glucose transport in perfused rat hearts

Additions of Hormones or Fatty Acid to Perfusate	Peak Ventricular Pressure	Intracellular Water Equilibriated with 3-O-methyl Glucose in 7 min
	mm Hg	%
Aerobic perfusions		
None	60	17
Palmitate	60	11
None	100	64
Palmitate	100	21
Insulin	60	76
Insulin + palmitate	60	53
Anaerobic perfusions		
None	60	59
Palmitate	60	82

From Neely et al , 1969

The molecular mechanisms involved in control of transport are unknown The action of insulin is thought to involve binding of the hormone to the outer surface of the cell membrane where it modifies the membrane components in such a way that the glucose-carrier complex can move through the membrane at a faster rate Acceleration of glucose transport by hypoxia or muscular work and inhibition of the process when fatty acids are oxidized are thought to be mediated by changes in intracellular levels of specific activators or inhibitors of transport Muscular work not only accelerates glucose transport in the absence of insulin but renders the transport process more sensitive to suboptimal concentrations of the hormone This combination of a direct acceleration by muscular work and increased sensitivity to insulin may explain the improved utilization of glucose by diabetics during exercise

Once glucose is within the cell, it must be

phosphorylated before it can be utilized further This reaction is catalyzed by hexokinase An important physiological property of this enzyme is inhibition by glucose-6-P (Crane and Sols, 1953) When the supply of glucose-6-P is more than adequate to meet the needs of the cell, the level increases and regulates its own production by inhibiting hexokinase The relationship between the rate of glucose phosphorylation and intracellular glucose-6-P concentration is shown in Figure 7 91

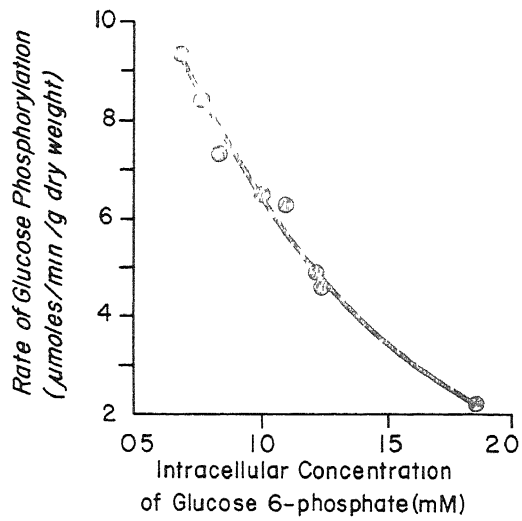
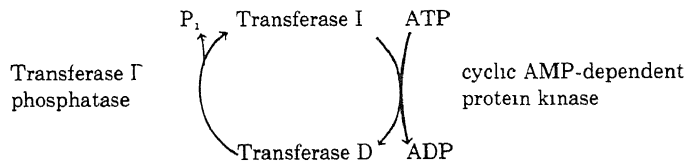


Fig 7 91 Relationship between intracellular glucose 6-phosphate concentration and rate of glucose phosphorylation in perfused rat heart (From England and Randle 1967)

GLYCOGEN HOMEOSTATIC UNIT

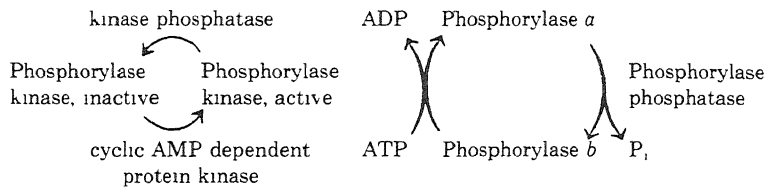
Glucose-6-P occupies the first branch point in glucose metabolism For glycogen synthesis it is first converted to glucose-1-P which combines with uridine triphosphate (UTP) to form uridine diphosphoglucose (UDPG) (Fig 7 88) The glucosyl unit of UDPG is transferred to glycogen by UDPG-glycogen transferase This enzyme is the site of regulation of glycogen synthesis by both hormonal and nonhormonal factors Transferase is a complicated enzyme which exists in two forms, transferase D and transferase I The D form is dependent upon the presence of glucose-6-P for activity while the I form has activity independent of this regulator The activity of the D form is inhibited by ATP so that variations in tissue levels of ATP and glucose-6-P increase or decrease the activity of this form Another important mechanism for control of transferase activity is interconversion of the two forms of the enzyme, as follows



Transferase I is phosphorylated to the D form by a protein kinase that is dependent upon cyclic adenosine monophosphate (cyclic AMP) Conversion of D to I involves dephosphorylation of the D form by a specific phosphatase These second-stage interconverting enzymes are the sites of much of the hormonal and nonhormonal control of glycogen synthesis Epinephrine causes net glycogenolysis by increasing intracellular cyclic AMP which activates conversion of transferase I to D and, as will be discussed, conversion of phosphorylase *b* to *a* Insulin increases glycogen synthesis by two mechanisms *1*) Conversion of transferase to the I form appears to be an effect of the hormone on protein kinase to increase its dependence on cyclic AMP, and thus to decrease its activity Increased levels of transferase I result (Fig 7 92) *2*) Insulin accelerates membrane transport and provides more substrate for hexokinase which increases the level of glucose-6-P The higher levels of glucose-6-P increase the activity of transferase D and, as will be discussed, reduce the activity of phosphorylase *b*

The net effect of insulin in the intact tissue is to increase glycogen levels The second-stage interconverting enzymes are also under nonhormonal control Increased levels of glycogen inhibit the phosphatase and reduce the amount of transferase in the I form (Fig 7 93) These changes decrease the rate of glycogen synthesis and set an upper limit on the level of glycogen that the cell can attain

Glycogenolysis is catalyzed by glycogen phosphorylase This enzyme, like UDPG-glycogen transferase, exists in two forms, phosphorylase *a* and *b*



Phosphorylase *b* is the dephosphorylated form and is active only in the presence of 5'AMP The *b* form is phosphorylated by a specific kinase enzyme, that requires Ca^{++}, and is converted to phosphorylase *a* The *a* form is active in the absence of 5'AMP Phosphorylase, like transferase, is subject to both hormonal and nonhormonal control

Many of the regulatory enzymes that exist in two forms have been called active and inactive In general, this is an unfortunate misnomer because both forms of the enzymes are capable of catalyzing their respective reactions The term "active" really means that this form of the enzyme is independent of control by allosteric regulators The so-called inactive form is inactive under some conditions that exist in cells because of the presence of allosteric inhibitors For example, phosphorylase *b* is activated by 5'AMP and is capable of as much activity per unit of enzyme protein as phosphorylase *a* This activity is not expressed in the tissue, however, because of high levels of inhibitors (ATP and glucose-6-P) normally found in cells With decreased levels of these inhibitors or increased levels of 5'AMP, phosphorylase *b* is activated (Fig 7 94) Variations in the tissue levels of ATP, AMP and glucose-6-P can control the rate of glycogenolysis by their action on phosphorylase *b* without interconversion of the enzyme forms Under most conditions phosphorylase is present in the cell predominantly in the *b* form and this mechanism appears to be important in regulating the utilization of glycogen The relationship between the tissue levels of glucose-6-P and the rate of glycogen utilization in heart muscle is shown in Figure 7 95 The tissue level of glucose-6-P was altered by perfusing hearts with various substrates When the level of glucose-6-P is low, glycogen breakdown is fast and, when glucose-6-P is high, net glycogen synthesis occurs This relationship between glucose-6-P and glycogen is due to effects of glucose-6-P on both activation of UDPG-glycogen transferase and inhibition of phosphorylase b

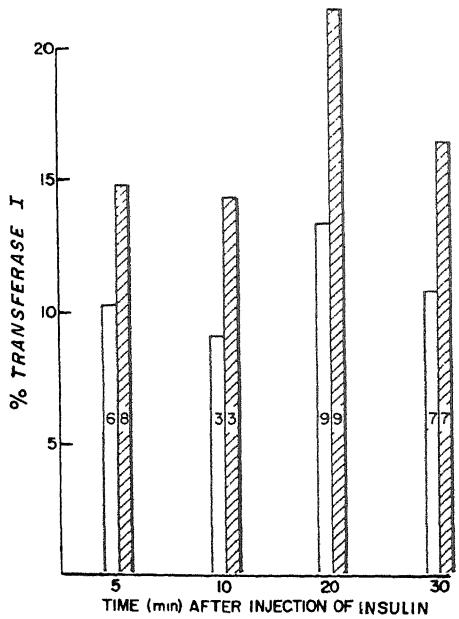


Fig 7 92 Per cent transferase I activity in extracts of rat skeletal muscle after intraperitoneal injection of saline (unshaded bars) or insulin (shaded bars) The numbers of experiments in each group are indicated (From Larner et al , 1968)

Interconversion of phosphorylase *a* and *b* is also an important physiological regulatory mechanism The activity of phosphorylase *b* is expressed as a defense against low energy levels in the cell only after the fact In contrast, phosphorylase activity and glycogen utilization increases in anticipation of energy requirements by converting the enzyme to the *a* form Phosphorylase *a*, the so-called active form is not subject to allosteric regulation as is the *b* form (Fig 7 94) By conversion to the *a* form, phosphorylase activity can be increased even in the presence of high levels of ATP and glucose-6-P and before there is an actual energy deficit Epinephrine is released into the blood in anticipation of muscular activity (flight-or-fight response) and this hormone promotes glycogenolysis by bringing about conversion of phosphorylase to the *a* form This action of epinephrine is mediated through increased levels of cyclic AMP in the tissue as indicated by the temporal sequence of changes in cyclic AMP and phosphorylase activity (Fig

7 96) The second-stage interconverting enzyme, phosphorylase kinase, also exists in two forms, so-called active and inactive A third-stage interconverting enzyme, protein kinase, phosphorylates inactive phosphorylase kinase and in so doing converts it to the active form This third-stage enzyme is activated by cyclic AMP With epinephrine treatment, the level of cyclic AMP is elevated and protein kinase is activated This enzyme then converts inactive phosphorylase kinase to the active form Phosphorylaese kinase catalyzes the phosphorylation of phosphorylase *b* and converts it to the *a* form which in turn brings about an increased rate of glycogenolysis

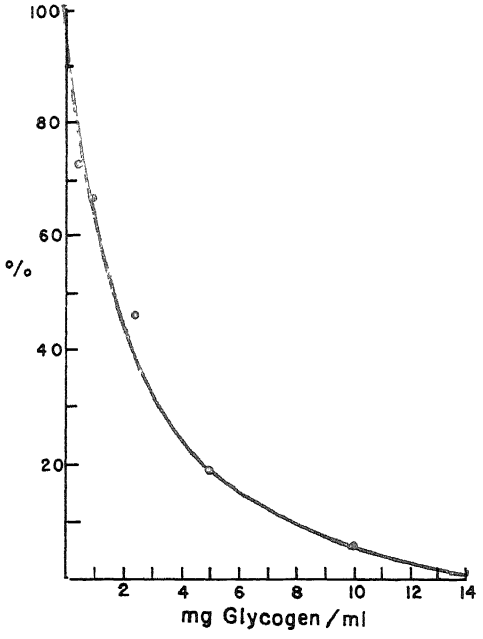


Fig 7 93 The inhibition of transferase phosphatase by glycogen (From Larner et al 1968)

Muscular contraction itself caused an increased conversion of phosphorylase to the *a* form in skeletal muscle In Figure 7 97 is shown a time course of phosphorylase conversion to the *a* form after onset of muscular contraction Two factors may account for this conversion of phosphorylase *b* to *a* during muscular contraction *1*) Phosphorylase *b* kinase is activated by calcium and depolarization of the muscle cell membrane releases calcium into the sarcoplasm (Krebs et al, 1968) The so-called inactive form of phosphorylase *b* kinase is activated by magnesium There is very little free magnesium in the cell and most of the ATP is in the form of ATP-Mg^{++} complex With ATP breakdown during muscular contraction, the free magnesium released from the ATP-Mg^{++} complex may activate phosphorylase *b* kinase (Villar-Palasi and Wei, 1970)

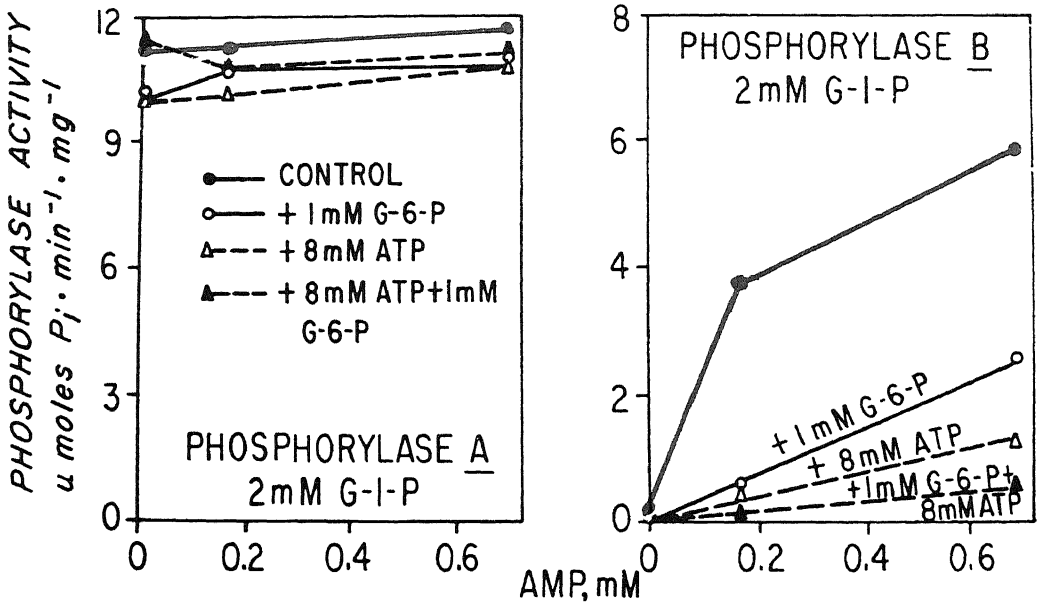


Fig 7 94 Effect of ATP and glucose-6-P on activity of phosphorylases *a* and *b* at various levels of AMP Assays were performed in the direction of glycogen synthesis by measuring release of inorganic phosphate (From Morgan and Parmegiani 1964)

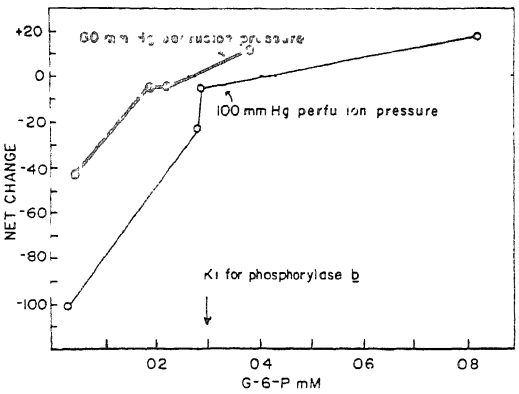


Fig 7 95 Relationship between net glycogen change and glucose 6 P concentration in perfused rat heart (From Neely et al 1968)

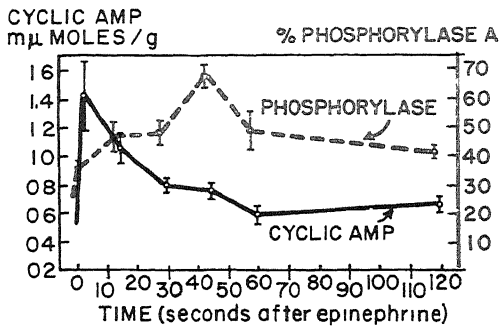


Fig 7 96 Effect of epinephrine on cyclic AMP concentration and phosphorylase activity in isolated rat hearts (From Robison et al 1965)

CONTROL OF THE GLYCOLYTIC UNIT

Phosphofructokinase catalyzes the first irreversible step in the glycolytic unit (glucose-6-P to acetyl CoA) Phosphofructokinase is a classic example of a regulatory enzyme and its activity controls the overall rate of glycolysis under many conditions The enzyme is subject to allosteric regulation by a variety of small molecules Its activity is inhibited by ATP and citrate (Fig 7 98) and is accelerated by P_i, 5'AMP, adenosine diphosphate (ADP), and cyclic AMP (Fig 7 99) This is the primary site of glycolytic inhibition by aerobic respiration When oxidative metabolism is proceeding normally, the tissue level of ATP is high and the levels of ADP, AMP, and P_i are low and phosphofructokinase is inhibited Inhibition of glycolysis by oxidation of fatty acids involves high tissue levels of ATP and citrate and low levels of AMP and P_i When the oxygen supply to muscle is limited, phosphofructokinase is activated by decreased levels of ATP and increased levels of 5'AMP and P_i

Pyruvate dehydrogenase catalyzes the conversion of pyruvate to acetyl CoA in the mitochondria Pyruvate dehydrogenase is a complex enzyme which is subject to allosteric control Activity of the enzyme is inhibited by high levels of NADH and acetyl CoA, both of which are products of the reaction (Randle et al, 1966) The enzyme is also inhibited by

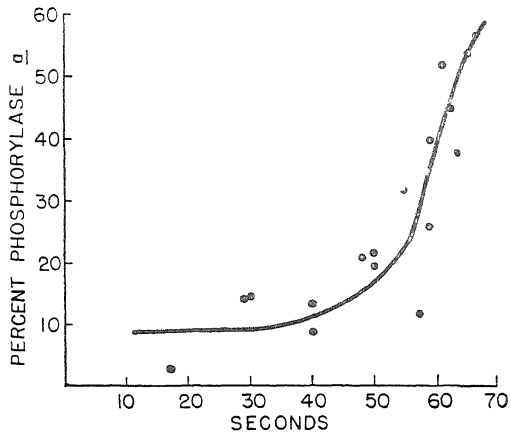


Fig 7 97 Time course of appearance of phosphorylase *a* in response to repetitive stimulation of frog sartorii at 20° C The muscles were stimulated 6 times per second with 15 volt shocks of 2 msec duration to contract against a 2-g preload The series of contractions was halted by freezing at the indicated time (From Danforth and Helmreich 1964)

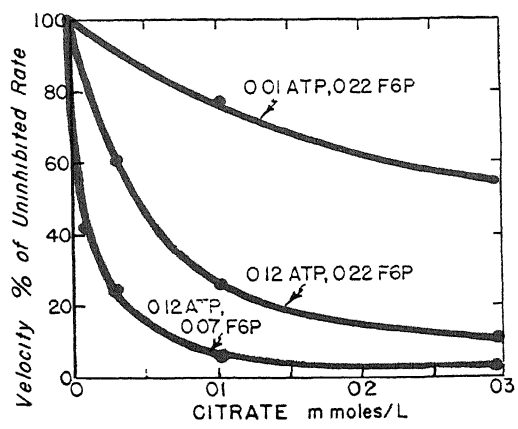


Fig 7 98 Velocity of phosphofructokinase as a function of citrate concentration at several ATP and fructose-6-P levels (From Passonneau and Lowry 1963)

ATP, the end product of pyruvate oxidation β-Oxidation of fatty acids is the major source of acetyl CoA in the mitochondria and, therefore, supplies a direct inhibitor of pyruvate dehydrogenase Inhibition by acetyl CoA and NADH has an important function in anaerobic muscle If oxygen supply is limiting, acetyl CoA cannot be used by the citric acid cycle and its production from pyruvate is inhibited NADH cannot be oxidized in the anaerobic tissue and its level builds up Inhibition of pyruvate dehydrogenase results in a higher level of pyruvate which is converted to lactate by lactic dehydrogenase This conversion of pyruvate to lactate reoxidizes the NADH produced in glycolysis and allows glycolysis to continue with anaerobic production of ATP

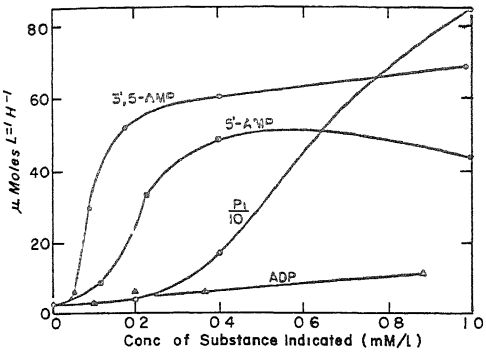


Fig 7 99 Effect of P_i 5′ AMP ADP and cyclic AMP on the activity of phosphofructokinase The other concentrations of reactants were ATP 2 3 mM Mg 5mM fructose-6-P 0 1mM The P concentrations as indicated are drawn to a 10-fold smaller scale (From Passonneau and Lowry 1962)

Pyruvate dehydrogenase, like UDPG-glycogen transferase and phosphorylase, is regulated by a phosphorylation-dephosphorylation mechanism (Reid, 1969) Phosphorylation, which is catalyzed by pyruvate dehydrogenase kinase, inactivates the enzyme while dephosphorylation by pyruvate dehydrogenase phosphatase activates the enzyme Activation of pyruvate dehydrogenase from pig heart is stimulated by cyclic AMP (Wieland and Siess, 1970) Pyruvate dehydrogenase and pyruvate metabolism may, therefore, be under hormonal control by interconversion of two forms of the enzyme

Table 7 12
Effects of various physiological states on glucose utilization by heart muscle[a]

Physiological State	Glucose-Activation Unit	Glycogen-Homeostatic Unit	Glycolytic Unit
Fasting, uncontrolled diabetes, hypophysectomized diabetic animal treated with growth hormone and cortisone	Slow due to low plasma insulin and high free fatty acids, impaired sensitivity to insulin	Stores glycogen due to high glucose-6-P levels resulting from inhibition of the glycolytic unit at phosphofructokinase	Inhibited at phosphofructokinase and pyruvate dehydrogenase by high levels of citrate and acetyl CoA
Hypophysectomized-diabetic animal	Slow due to low plasma insulin but insulin sensitivity is increased	Uses glycogen due to low glucose-6-P levels	Lower levels of citrate and acetyl CoA allow glycolysis to proceed
Increased heart work	Fast due to effect of heart work on transport and increased insulin sensitivity These effects may be blocked if free fatty acid levels are high	Uses glycogen due to low glucose-6-P levels which result from activation of the glycolytic unit	Activated by lower tissue levels of citrate and acetyl CoA and higher levels of P_i
Hypoxia resulting from coronary insufficiency	Fast due to stimulation of glucose transport and to increased sensitivity to insulin	Uses glycogen due to lower ATP and glucose-6-P levels and to higher levels of AMP and P_i	Activated by lower tissue levels and ATP and higher levels of P_i and 5′AMP

[a] Changes are related to activities of the various units in the postabsorptive state

Integrated Control of Carbohydrate Metabolism

Regulation of the glucose-activation, glycogen-homeostatic, and glycolytic units is integrated to allow for production of ATP as needed but to prevent useless consumption of glucose. As summarized in Table 7.12, fasting, uncontrolled diabetes, and hormone-treated hypophysectomized diabetic animals are characterized by slow rates of glucose utilization resulting from low plasma levels of insulin and high levels of free fatty acid. Inhibition of both the glycolytic and glucose-activation units is present. The heart stores much of the glucose that is taken up as glycogen due to high tissue levels of glucose-6-P. Plasma of hypophysectomized diabetic animals does not contain as much free fatty acid so that both glucose activation and glycolysis can proceed at a faster rate. These animals are characterized by large variations in the rate of glucose utilization due to greatly increased sensitivity of transport to stimulation by insulin. Increased heart work and hypoxia are characterized by activation of all three units. The work effect can be blocked by high levels of fatty acids but the effect of hypoxia cannot.

Regulation of carbohydrate metabolism in skeletal muscle is similar to regulation in heart muscle in most respects. The major difference is in the extent to which fatty acid oxidation is able to inhibit the glucose-activation and glycolytic units. In heart muscle, fatty acid is a powerful inhibitor while in skeletal muscle, oxidation of this substrate does not account for as large a percentage of oxygen consumption and does not result in as profound an inhibition of carbohydrate metabolism. As a consequence, skeletal muscle of fasting or diabetic animals uses more glucose and stores less glycogen than heart muscle.

11 / CONTROL OF FATTY ACID METABOLISM IN ADIPOSE TISSUE

Adipose tissue contains the largest store of fuel of any tissue in the body The average person carries a 40-day reserve which is stored as triglycerides (Dole, 1965) The major physiological function of adipose tissue is maintenance of blood free fatty acid (FFA) levels to insure a constant supply of oxidizable substrate This function involves *1*) removal of fatty acids and their precursors from plasma, 2) storage of these fuels as triglycerides during periods of substrate excess, and *3*) breakdown of triglycerides with release of FFA into the plasma during periods of increased energy demands or deficiency of exogenous substrates The major metabolic pathways that are involved in fatty acid metabolism in adipose tissue are outlined in Figure 7 100 Most of the conversion of glucose to fatty acids that occurs in the body occurs in this tissue Fatty acids which have been synthesized by the liver or absorbed by the intestine are also taken up and stored as triglyceride Fatty acids are present in plasma either as the FFA-albumin complex or as triglycerides attached to lipoproteins or present in chylomicrons Lipoprotein lipase hydrolyzes serum triglycerides at the capillary membrane to give FFA and glycerol Upon appropriate metabolic or hormonal signals, stored triglyceride which is present in the fat cell can be hydrolyzed and the fatty acid and glycerol returned to the plasma

Storage of excess calories in the form of triglyceride rather than carbohydrate or protein has several advantages Oxidation of protein or carbohydrate yields about 4 kcal /g , whereas oxidation of fatty acids yields 9 kcal /g Storage of protein or carbohydrate also requires accumulation of intracellular water while triglyceride may comprise as much as 90% of the net weight of adipose tissue (Cahill et al , 1970) In terms of tissue accumulated per calorie, there is about 1 g /kcal for carbohydrate or protein and about 0 1 g /kcal for fat This difference in caloric content is reflected in body weight loss during fasting For the 1st day or two a person uses more carbohydrate and protein and looses weight rapidly, but upon extended fasting a larger portion of the metabolic needs are derived from fatty acids and the rate of weight loss decreases During even the early stages of fasting, however, a major part of the CO_2 produced by the body comes from oxidation of fatty acids directly or from oxidation of ketone bodies that are derived from fatty acids (Fig 7 101) As starvation progress, all of the substrate oxidized by carcass and viscera comes from fatty acids and ketoacids while about ⅔ of the substrate used by brain is derived from ketoacids (Fig 7 102)

This chapter will be concerned with the regulation of fatty acid synthesis, storage, and mobilization in adipose tissue by hormonal and nonhormonal mechanisms For a detailed discussion of this topic, the student is referred to two recent publications (Renold and Cahill, 1965, Jeanrenaud and Hepp, 1970)

REGULATION OF FATTY ACID SYNTHESIS

Since the function of the adipocyte is storage and mobilization of lipids, metabolic activity of the cell is specially adapted for this function and consequently some regulatory mechanisms are more pronounced or even unique to this tissue Glucose is the precursor of greatest physiological importance for lipid synthesis It is used as a source of acetyl CoA, energy and reducing power for fatty acid synthesis, and as a source of α-glycerol-P for fatty acid esterification (Fig 7 100) Insulin plays a dominant role in regulating lipogenesis by its action on glucose transport Low levels of insulin increase the conversion of ^{14}C-glucose to fatty acids (Fig 7 103) This effect of insulin is

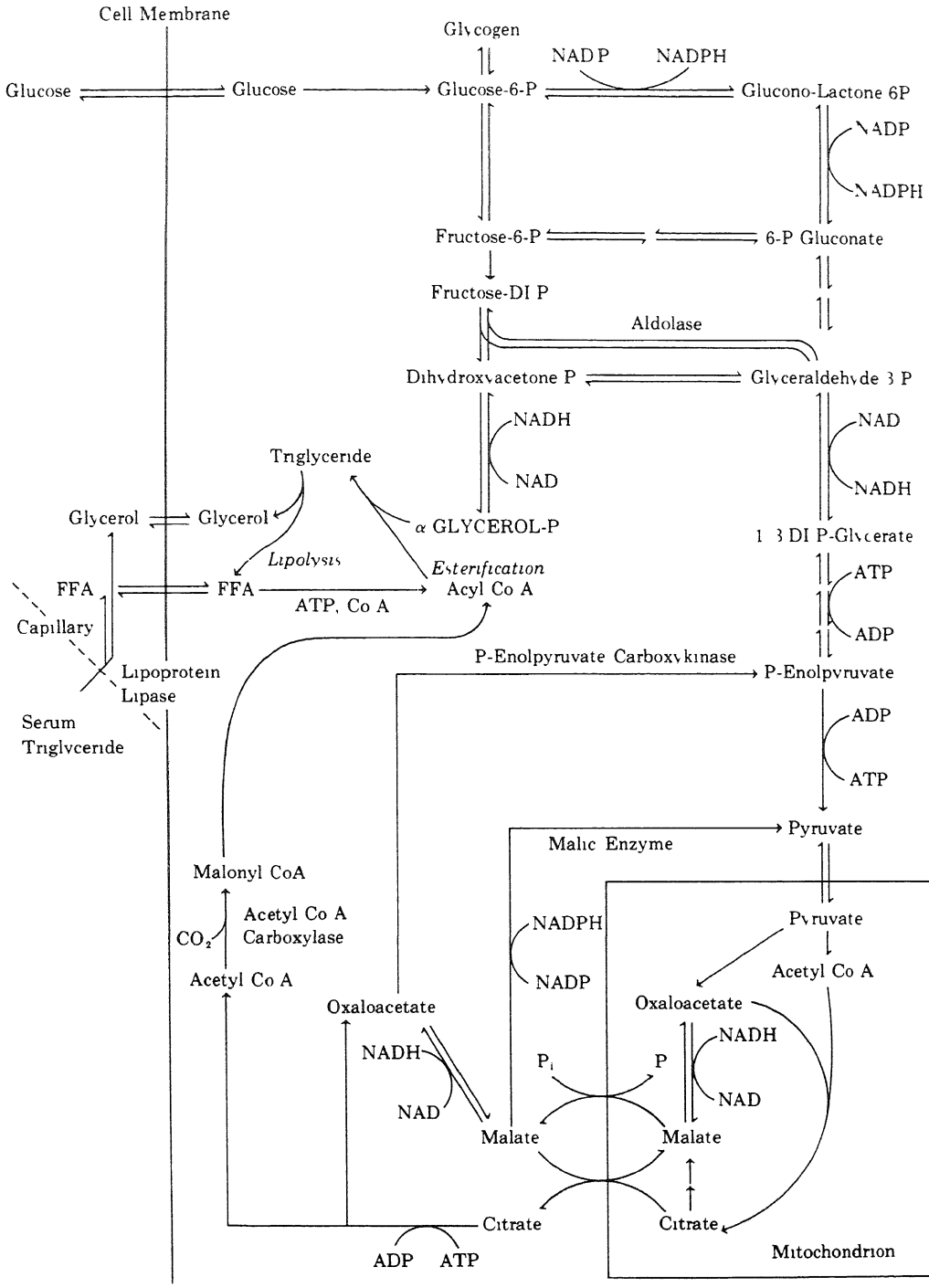


Fig 7 100 Pathways regulating fatty acid metabolism in adipose tissue

associated with an increased glucose uptake The effect of insulin on fatty acid synthesis is greatly reduced in fasting animals The mechanism of the reduction will be discussed presently Other steps in glucose utilization may be controlled by hormones For example, epinephrine increases glucose uptake in conjunction with activation of lipolysis This effect

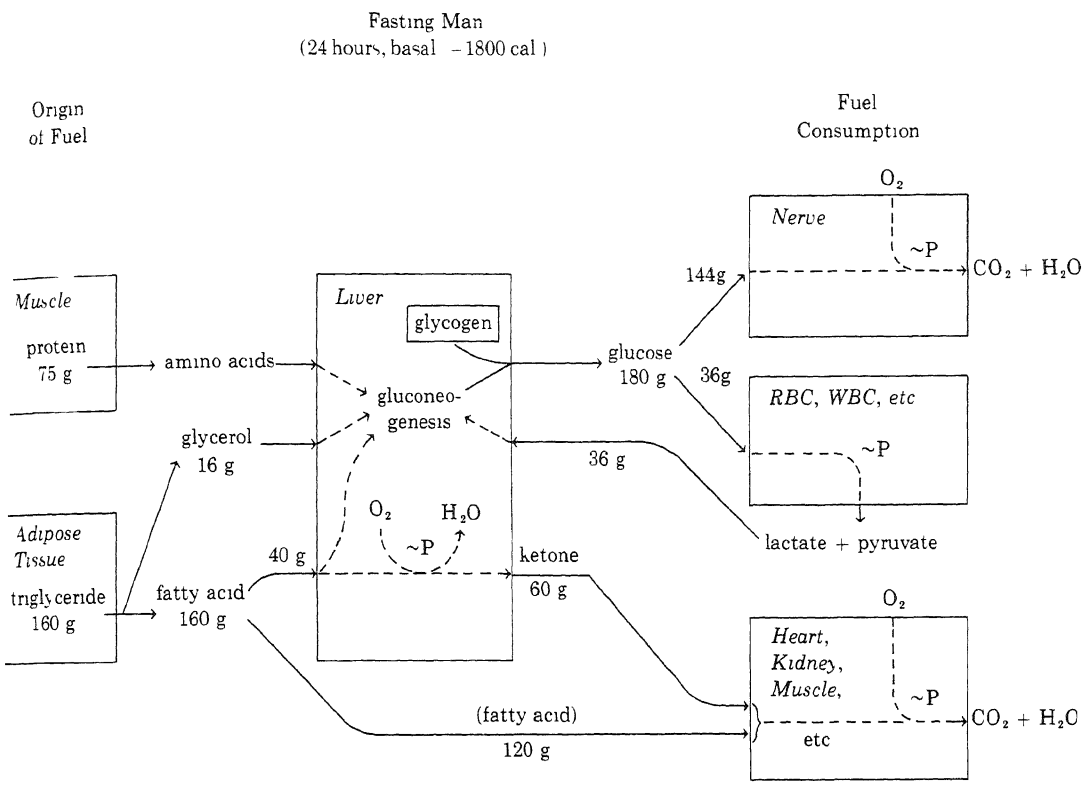


Fig 7 101 Generalized scheme of daily fuel metabolism in a fasting man (From Cahill et al 1970)

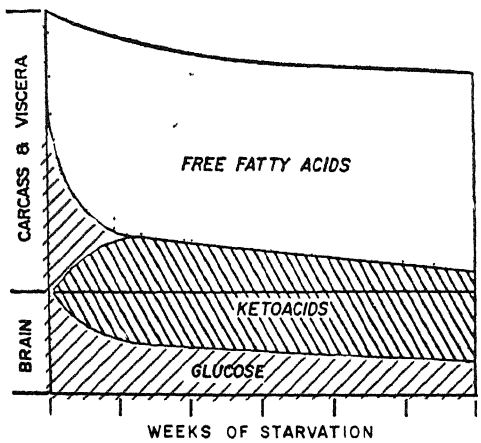


Fig 7 102 Substrate oxidation in fasting man Schematic representation of the transition from glucose to fatty acid utilization by the carcass and viscera and from glucose to ketoacids by the brain (From Cahill et al 1970)

involves activation of phosphofructokinase by decreased levels of adenosine triphosphate (ATP) and increased levels of cyclic adenosine monophosphate (cyclic AMP) This enzyme, like that discussed for muscle, is inhibited by ATP and citrate and is activated by cyclic AMP, AMP, adenosine diphosphate (ADP), and P_i (Denton and Martin, 1970)

Conversion of glucose to pyruvate occurs in the cytoplasm Pyruvate enters the mitochondria where it is converted to acetyl CoA Since fatty acid synthesis occurs in the cytoplasm, transport of acetyl CoA out of the mitochondria must occur (Fig 7 100) Mitochondrial membranes are impermeable to acetyl CoA which necessitates conversion of this lipid precursor to other compounds that can penetrate the membrane Conversion to acetyl carnitine, catalyzed by acetyl CoA-carnitine transferase, accounts for only a small per cent of acetyl transfer due to low activity of the enzyme The major mechanism for transport of acetate units involves conversion to citrate which is transported to the cytoplasm in exchange for malate (Fig 7 100) Citrate is converted to acetyl CoA and oxaloacetate in the cytoplasm by an ATP-dependent citrate lyase Transport of acetyl CoA from mitochondria to cytoplasm does not limit conversion of glucose to fatty acid (Flatt, 1970) This conclusion is based on the observation that addition of pyruvate in

the presence of glucose increases fatty acid synthesis indicating that the capacity of mitochondria to synthesize and export acetyl CoA is greater than the rate observed with glucose alone

Oxaloacetate arising from citrate cleavage must be transported back into the mitochondria to sustain citrate synthesis Conversion to malate in the cytoplasm, which is then exchanged for more citrate, serves to *1)* maintain the intramitochondrial pool of citric acid cycle intermediates, and *2)* to shuttle reducing power into the mitochondria When glucose is converted to fatty acids, production of reduced nicotinamide adenine dinucleotide (NADH) in the cytoplasm exceeds its utilization (Flatt, 1970) Excess reducing power is transferred to

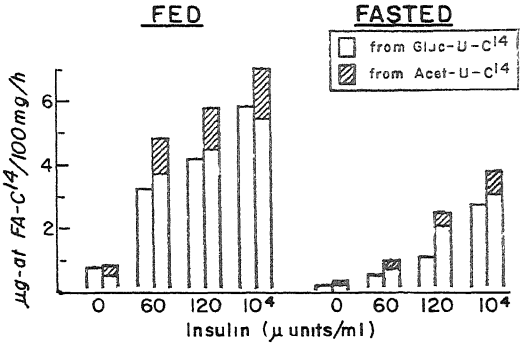


Fig 7 103 Incorporation of glucose-C and acetate-C into fatty acid by adipose tissue of fed rats or rats starved for 48 hr In each pair of columns open areas show the incorporation of glucose-C during incubation with glucose-U-^{14}C (left column) or with glucose-U-^{14}C in the presence of unlabeled acetate (right column) Incubation with unlabeled glucose and acetate-U-^{14}C in the medium gave the incorporation shown by hatched areas these are represented on the top of the right-hand column to show the total rate of fatty acid synthesis in the presence of the two substrates (From Flatt 1970)

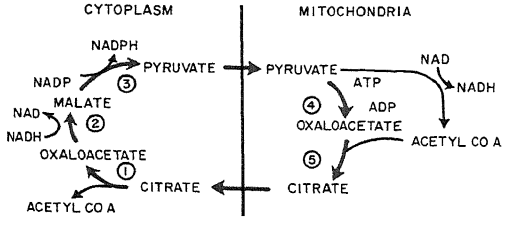


Fig 7 104 Malate cycle Enzymes involved in the cycle are indicated as follows *1)* citrate lyase *2)* malic dehydrogenase, *3)* malic enzyme *4)* pyruvate carboxylase- and *5)* citrate synthetase (Adapted from Flatt 1970)

the mitochondria by converting oxaloacetate to malate in the cytoplasm and malate back to oxaloacetate in the mitochondria (Fig 7 100) Only a small part of the oxaloacetate formed in the cytoplasm by citrate cleavage is needed, however, to transfer reducing power into the mitochondria (Denton and Martin, 1970) This means that there must be other mechanisms *1)* to get oxaloacetate into the mitochondria for citrate synthesis, and *2)* to get malate out of the mitochondria for countertransfer with citrate The most important pathway for transfer of oxaloacetate into the mitochondria is the "malate cycle" (Fig 7 104) Oxaloacetate produced in the cytoplasm is converted to malate by malic dehydrogenase Malate is then converted to pyruvate by NADP-malic dehydrogenase (malic enzyme) and pyruvate is carboxylated via pyruvate carboxylase to yield oxaloacetate in the mitochondria, thus completing a cycle This pathway results in transfer of reducing power from NADH to NADP in the cytoplasm where it can be used for fatty acid synthesis The activity of malic enzyme correlates with the rate of lipogenesis The activity is decreased during fasting or diabetes and is increased following insulin treatment Malate is transferred into the cytoplasm by a separate dicarboxylate carrier which involves exchange with phosphate (Fig 7 100)

Cytoplasmic acetyl CoA is converted to fatty acids by carboxylation to malonyl CoA Acetyl CoA carboxylase, the enzyme catalyzing this reaction, has been proposed as a regulator of fatty acid synthesis since under some conditions it appears to be controlled by tissue levels of citrate (an activator) and fatty acyl CoA (an inhibitor) Under most conditions where there is a change in the rate of fatty acid synthesis, no consistent relationship between the levels of these effectors and the rate of fatty acid synthesis can be found Regulation at this step may be important, however, during stimulation of fatty acid synthesis by insulin In this instance, triglyceride synthesis is stimulated through increased levels of α-glycerol-P and ATP which results in lower levels of fatty acyl CoA These changes may allow acetyl CoA carboxylase to proceed faster

Reducing power for fatty acid synthesis is supplied by two processes *1)* The pentose phosphate pathway produces two molecules of NADPH for each revolution of the cycle *2)* Conversion of malate to pyruvate in the

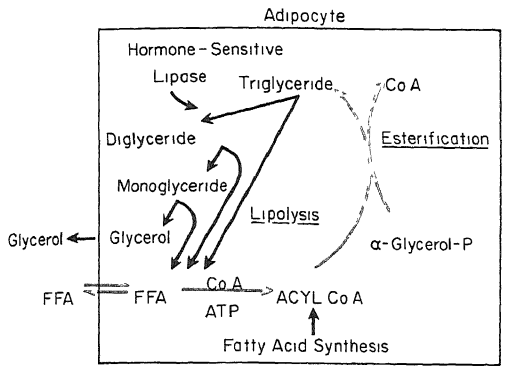


Fig 7 105 The triglyceride homeostatic unit

"malate cycle" produces one NADPH for each acetyl CoA transported out of the mitochondria (Fig 7 104) Capacity for producing cytoplasmic reducing power during conversion of glucose to fatty acid is more than is required for fatty acid synthesis (Flatt, 1970) Some of the reducing power can be used in the production of α-glycerol-P, but most of the NADH is available for ATP production by oxidative phosphorylation Unlike most synthetic pathways which utilize energy, conversion of glucose to fatty acids is an energy-generating process and with high levels of insulin, the maximal rate of fat synthesis from glucose may be limited by the tissue's ability to utilize the excess energy

TRIGLYCERIDE HOMEOSTASIS

Synthesis and hydrolysis of triglycerides involve separate synthetic and degradative pathways that are termed esterification and lipolysis, respectively The two pathways form a cyclic process of synthesis and breakdown which is in a dynamic state at all times (Fig 7 105) Net release or deposition of FFA depends on the relative rates of the two pathways Each triglyceride contains 3 molecules of fatty acid esterified to the hydroxyl groups of one glycerol molecule FFA must be activated to form fatty acyl CoA at the expense of ATP before esterification occurs The driving energy for esterification is furnished by hydrolysis of these high energy CoA bonds and of the high energy phosphate bond of α-glycerol-P Hydrolysis of triglyceride yields fatty acid and glycerol in a ratio of three to one Enzymatically, this process occurs in a steplike manner, one fatty acid being released at a time A separate enzyme is responsible for each step As a group, these enzymes are called lipases

Regulation of esterification occurs primarily by indirect mechanisms related to production of substrates α-glycerol-P is required for esterification and must be produced from glucose or pyruvate since glycerol kinase, the enzyme responsible for direct phosphorylation of glycerol, is not present in adipose tissue The action of insulin to increase esterification can be explained in part by increased levels of α-glycerol-P secondary to increased glucose transport and higher levels of dihydroxyacetone-P and NADH Another source of α-glycerol-P is by a process called "glyceroneogensis" by analogy to gluconeogenesis Adipose tissue has the capacity to carry out the reactions of the gluconeogenic pathway as far as production of fructose-di-P (Fig 7 100) Fructose diphosphatase is not present so that formation of glucose or glycogen cannot occur The series of reactions involved in glycereoneogenesis is as follows

$$\text{pyruvate} \rightarrow \text{oxaloacetate} \xrightarrow{\text{P enol pyruvate carboxykinase}} \rightarrow \rightarrow \rightarrow$$

$$\text{P-enol pyruvate} \rightarrow \text{dihydroxyacetone-P} \downarrow \alpha\text{-glycerol-P}$$

This pathway is probably utilized only during glucose deficiency such as starvation or diabetes and functions to protect the cell against uncontrolled breakdown of triglycerides (Table 7 13) Increased conversion of pyruvate to glycerol is accompanied by increased activity of P-enol pyruvate carboxykinase which indicates that this enzyme may be an important regulator of glyceroneogenesis Addition of pyruvate to adipose tissue, in vitro, reduces the rate of FFA release by increasing esterification which can be accounted for by increased formation of α-glycerol-P (Reshef et al , 1970) Since fructose diphosphatase is not present in adipose tissue, conversion of triose-P to fructose-di-P during glyceroneogenesis is unnecessary This conversion is prevented by a decline in the activity of aldolase during fasting or diabetes

Fatty acids for esterification are supplied from three main sources. *1*) synthesis from glucose, *2*) uptake from blood, and *3*) hydrolysis of stored triglyceride Regulation of the activity of lipoprotein lipase is involved in controlling the supply of fatty acids from the

Table 7 13
Glyceroneogenesis in adipose tissue in fasting and diabetes[a]

Treatment	Pyruvate Incorporation		Enzyme Activity		
	Fatty acids	Glycerol	Phosphoenol-pyruvate carboxykinase	Pyruvate carboxylase	NADP-malate dehydrogenase
	nmoles/min mg protein		*nmoles/min mg protein*		
Fed ad libitum	6 8 ± 1 0	0 9 ± 0 2	15 ± 2	32 ± 2	115 ± 11
Fasted, 24 hr	2 9 ± 1 0*	1 7 ± 0 2*	34 ± 5*	27 ± 5	94 ± 8
Alloxan diabetic	3 2 ± 0 9*	2 5 ± 0 3*	48 ± 13*	18 ± 3	55 ± 7*

From Shafrir et al 1970

*Pyruvate incorporation was measured by incubation of rat epididymal fat tissue for 2 hr in buffer containing ^{14}C-pyruvate To compare the incorporation values with the enzyme activities, all results are expressed in nmoles of pyruvate/min mg total tissue protein Values given are means ± S E Asterick denotes significant difference from control

blood (Robinson and Wing, 1970) This enzyme acts at the capillary endothelium to hydrolyze chylomicron and lipoprotein triglyceride to free fatty acids and glycerol (Fig 7 100) The activity is increased during incubation of adipose tissue from starved animals, which initially contains low enzyme activity (Fig 7 106) This increase can be prevented by including inhibitors of protein synthesis in the incubation If these inhibitors are added to tissue from fed rats which contain high levels of activity, a rapid decline in lipoprotein lipase occurs These findings indicate that synthesis of new enzyme or cofactor may be involved in this control mechanism In addition, insulin increases the activity of lipoprotein lipase, and epinephrine and other hormones that accelerate lipolysis inhibit the activity These hormone effects may involve changes in levels of cyclic AMP since addition of the nucleotide also inhibits lipoprotein lipase Thus, lipoprotein lipase appears to be controlled by *1)* rapid turnover of a protein component of the system, and *2)* regulation of the activity by cyclic AMP Activity is increased during periods of excess substrate supply which makes more FFA available for esterification and is decreased during starvation when lipolysis is stimulated in the fat cell

Hydrolysis of triglyceride is an important source of intracellular FFA Increased lipolysis stimulates esterification somewhat due to increased availability of fatty acyl CoA This phenomenon is seen when lipolysis is increased by epinephrine addition to fat cells Both glycerol and FFA are released, but glycerol release is relatively greater due to re-esterification of a

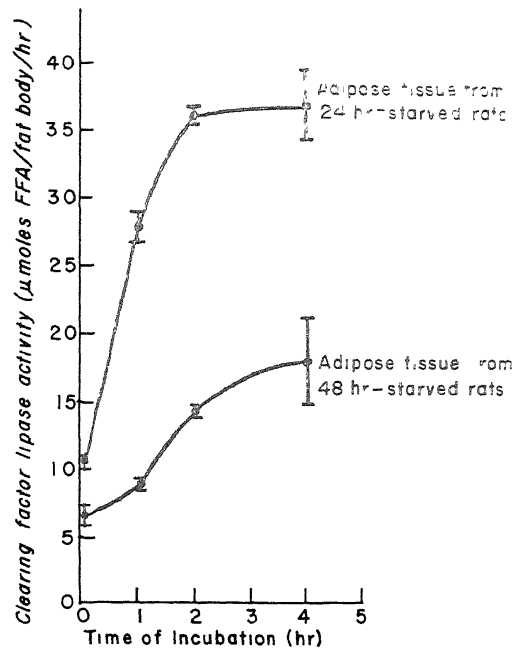


Fig 7 106 Rise in clearing factor lipase activity when adipose tissue from starved rats is incubated in vitro (From Robinson and Wing 1970)

portion of the FFA (Steinberg and Vaughn, 1965) This finding emphasizes the cyclical nature of the triglyceride homeostatic unit

Fatty acids that are supplied for esterification from the blood or from stored triglyceride must be activated to form fatty acyl CoA This reaction requires ATP and yields AMP and P-P Hormonal stimulation of lipolysis in fat cells is accompanied by a fall in ATP levels,

particularly when glucose is omitted from the incubation medium ATP levels increase markedly when glucose is added Therefore, availability of fatty acyl CoA may be regulated in two ways *1)* by control of the glycolytic and fatty acid synthetic pathways, and 2) by control of energy production within the cell

REGULATION OF LIPOLYSIS

Lipolysis can be regulated in two ways *1)* by an indirect mechanism involving changes in the production or removal of the products and substrates of the reactions, and *2)* by direct regulation of lipases FFA is an inhibitor of lipolysis Thus, dietary intake of fatty acid, increased synthesis from glucose, or hormonal activation of lipoprotein lipase will elevate intracellular levels of FFA and reduce the rate of lipolysis Direct hormonal regulation of lipolysis occurs at the level of the triglyceride lipase This is the rate-limiting enzyme in the lipolytic pathway since the hydrolytic rate from diglyceride to free glycerol is approximately 10 times faster than the rate from triglyceride to diglyceride Therefore, release of the first fatty acid is followed by a rapid complete hydrolysis to FFA and glycerol

Activity of triglyceride lipase is regulated by a large number of hormones (Table 7 14) For this reason, triglyceride lipase is referred to as "hormone-sensitive lipase " There are at least 12 hormones that are known to stimulate lipolysis in adipose tissue Epinephrine, norepinephrine, glucagon, adrenocorticotropic hormone (ACTH), secretin, thyroid-stimulating hormone (TSH), and luteinizing hormone (LH) are examples of lipolytic hormones The action of these hormones is mediated by stimulation of adenyl cyclase which results in increased levels of cyclic AMP The effect of ACTH and glucagon on cyclic AMP levels and the antagonism of this effect by insulin are shown in Figure 7 107 The contention that lipolytic hormones act through increased levels of cyclic AMP is supported by the direct effect of cyclic AMP to mimic the hormonal effect In addition, agents which inhibit phosphodiesterase, the enzyme which degrades cyclic AMP, increase intracellular cyclic levels and intensify the hormonal effect The lipolytic response to hormones and cyclic AMP reaches maximum within a few minutes. In addition to activation of triglyceride lipase, lipolytic hormones accelerate glucose transport and ion fluxes across the membrane

Table 7 14
Activation of lipase activity by pretreatment of fat pads with lipolytic hormones[a]

Hormone	*n*	Lipase Activity (μmoles glycerol liberated/60 min $\times$ g fresh wt)
None	18	2 49 ± 0 84
Epinephrine 10 μg/ml	24	3 98 ± 0 81
Glucagon 4 μg/ml	12	4 34 ± 0 75
ACTH 0 2 U/ml	7	7 6 ± 0 4
TSH 10 μg/ml	3	5 8 ± 1 0

From Loffler and Weiss, 1970

[a] Lipase activity was estimated by measuring release of glycerol

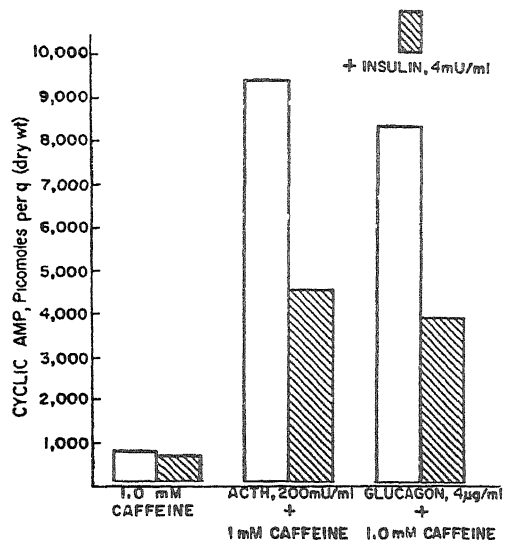
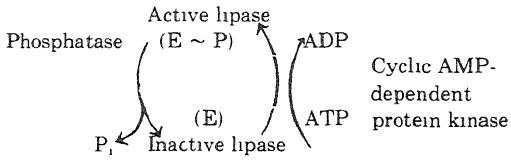


Fig 7 107 Effects of insulin on cyclic AMP levels in fat cells incubated with ACTH or glucagon Caffeine was added to minimize destruction of cyclic AMP by the phosphodiesterase (From Butcher et al 1968)

Hormone-sensitive lipase, like many other regulated enzymes, exists in active and inactive forms Activation involves phosphorylation of the inactive form that is catalyzed by a protein kinase which is dependent upon cyclic AMP (Corbin and Krebs, 1969, Hultunen et al , 1970) When activation occurs, radioactive phosphate from ATP is incorporated into the

lipase (Fig 7 108) The time course of phosphorylation parallels that for activation The following cycle of activation and inactivation is proposed for the hormone-sensitive lipase



The close similarity of this cycle to the cycle of reactions involved in control of phosphorylase activity (chap 10) should be apparent

Growth hormone and glucocorticoids also increase lipolysis but the mechanisms of action are quite different from those described for epinephrine and glucagon Growth hormone stimulates cyclic AMP production in isolated fat cells after a lag of at least 1 hr The hormone effect can be blocked by inhibitors of ribonucleic acid (RNA) and protein synthesis, indicating that new enzyme formation is involved The effect of growth hormone is increased by addition of glucocorticoids In addition to potentiating the effects of growth hormone, glucocorticoids play a "permissive" role in the effects of epinephrine, ACTH, and other hormones on lipolysis

The major antilipolytic hormone is insulin In addition to its effect on glucose transport which results in stimulation of lipogenesis, the hormone directly inhibits production of free fatty acid and glycerol in adipose tissue, even

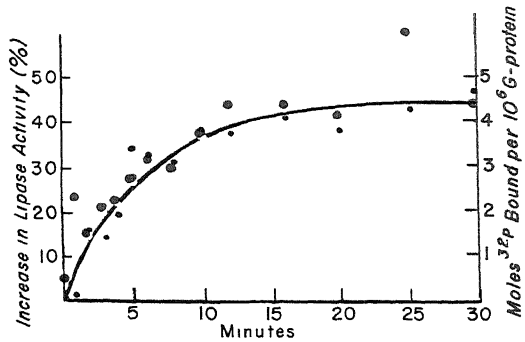


Fig 7 108 Time course of activation and phosphorylation of highly purified lipase in the presence of rabbit muscle protein kinase Aliquots were removed for determinations of lipase activity and protein-bound ^{32}P at indicated time intervals (From Hultunen et al 1970)

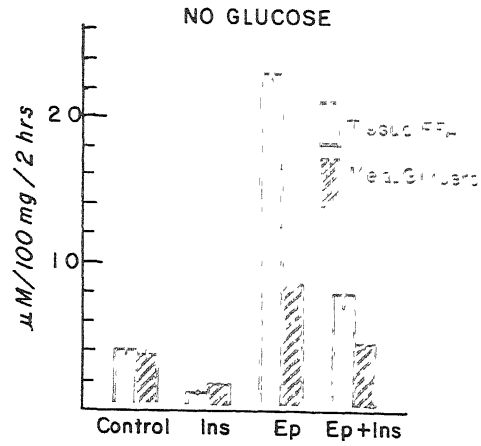


Fig 7 109 Effects of epinephrine and insulin on free fatty acid and glycerol production of adipose tissue (From Jungas and Ball 1963)

in the absence of glucose (Fig 7 109) Appearance of FFA and glycerol occurs at a low rate during control incubations, but this rate is inhibited about 50% by insulin Epinephrine accelerates production of FFA and glycerol and insulin inhibits appearance of both products This antilipolytic effect of insulin is mediated through a decrease in levels of cyclic AMP (Fig 7 107) As a result of reducing cyclic AMP levels, insulin also stimulates lipoprotein lipase The net effect, therefore, is an increase in fat deposition that is accomplished by sparing of triglyceride stores, by accelerating incorporation of plasma triglyceride into cell stores, and by accelerating conversion of glucose into fat

INTEGRATED CONTROL OF FAT METABOLISM

Control of fatty acid metabolism in adipose tissue serves to maintain a constant supply of oxidizable substrate in the plasma The major hormones involved in these controls include insulin, growth hormone, glucocorticoids, and catecholamines Net fatty acid storage or release reflects the balance between esterification and lipolysis (Table 7 15) Esterification is rapid when insulin, glucose and fatty acid are present in plasma in optimal amounts These rapid rates result from *1)* conversion of glucose to fatty acid, *2)* an optimal supply of α-glycerol-P, and *3)* increased lipoprotein lipase activity which facilitates uptake of lipid from the

Table 7 15
Integrated control of fat metabolism during fasting feeding and diabetes[a]

Physiologic States	Plasma Levels of Hormones and Substrates	Lipolysis	Esterification
Fasting, uncontrolled diabetes, hypox diabetic treated with growth hormone and cortisone	Insulin levels are low fatty acid levels are high, plasma triglycerides are high, ketone bodies are high	Stimulated due to release of insulin inhibition, together with higher levels of steroids and growth hormone Cyclic AMP levels increase	Slow due to low rate of glycolysis and fatty acid synthesis α-glycerol-P levels are low and endogenous fatty acid oxidation occurs to meet energy demands
Hypophysectomized diabetic	Insulin, ACTH, cortisone, and growth hormone levels are low Glucose levels may be high	Slower than uncontrolled diabetic due to release of lipase stimulation by pituitary and adrenal cortical hormones	Slow due to low plasma insulin Glucose conversion to fat impaired
Caloric excess	High glucose, fat, and insulin	Slow due to antagonistic effects of insulin and FFA	Fast due to increased glucose uptake resulting in conversion to α-glycerol-P and fatty acid Fatty acid taken up and esterified

[a] Changes are related to conditions in normal, fed animals

plasma Under these circumstances, lipolysis is inhibited by the antagonistic effects of insulin During starvation, insulin levels are low and growth hormone levels are elevated As a result, lipolysis is more rapid than esterification Large changes in the rates of lipolysis and esterification occur in starvation, diabetes, feeding, and in other hormone-deficiency states

12 CONTROL OF GLUCONEOGENESIS, KETOGENESIS, AND LIPOGENESIS IN

Gluconeogenesis

The central nervous system, the red blood cells, and possibly other tissues depend almost entirely on glucose as an energy source Maintenance of adequate glucose levels in the blood, on which the life of the animal depends during starvation, requires emergency and long-term adaptation systems Blood glucose is maintained during the postabsorptive state first by the breakdown of liver glycogen (glycogenolysis) and secondly by the process of glucose formation from nonglucose precursors This process is termed gluconeogenesis and is quantitatively the more important The rate of glucose production by gluconeogenesis can vary from 0 to above 300 g /24 hr in the adult human subject (Table 7 16) Only liver and kidney tissue appear to have the enzymatic capacity to carry out gluconeogenesis (Krebs, 1964, Scrutton and Utter, 1968) Based on the difference in organ weight and assuming similar overall gluconeogenic capacity in both tissues, the liver would be expected to contribute 90 to 95% of the total glucose production by the body in the postabsorptive state

GLUCONEOGENIC PATHWAY

A diagram of the gluconeogenic pathway is shown in Figure 7 110. To gain a clearer definition of the important features of the pathway, the relationship of glucose synthesis to other metabolic processes is not indicated, although such interactions may be important in the regulation of gluconeogenesis as will be discussed later Glucose can be formed from a variety of precursors which enter the gluconeogenic pathway at different points The pathway represents the major means of removing from the blood excessive quantities of lactate produced by the muscles during exercise and glycerol released from adipose tissue during fat mobilization It also represents an important catabolic pathway for many of the amino acids

Gluconeogenic tissues appear to be unique in that they possess the enzymatic potential for both glucose degradation via the glycolytic pathway and glucose synthesis Although the rate of glycolysis in these tissues is assumed to be low so that there is little impedance to the synthesis of glucose, necessity for the control of the direction of glucose metabolism is obvious The conversion of pyruvate to glucose is catalyzed by several of the enzymes of the glycolytic pathway and represents, therefore, a reversal of glycolysis There are three steps in the glycolytic pathway which are irreversible, however, and therefore different enzymes are required in order to bypass these steps in tissues that have gluconeogenic capacity The irreversible steps often represent the rate-limiting enzymes in metabolic pathways and provide key sites for regulation of glycolytic and gluconeogenic flux The thermodynamically irreversible steps of glycolysis and gluconeogenesis are shown in Figure 7 111 The enzymes that catalyze these steps in glucose synthesis from pyruvate are pyruvate carboxylase, phosphopyruvate carboxykinase, fructose-1, 6-diphosphatase and glucose-6-phosphatase Their activity is either very low or completely absent in all tissues except liver and kidney It can be seen in Figure 7 110 that the portion of the gluconeogenic pathway between pyruvate and phosphopyruvate represents a special problem since it involves both intra- and extramitochondrial reactions The enzyme pyruvate carboxylase appears to be compartmentalized within the mitochondrion Pyruvate must therefore enter the mitochondrion, perhaps by a carrier-mediated process, where it is carboxylated to oxalacetate by pyruvate carboxylase Oxalacetate must then be transferred out of the mitochondrion, since the enzyme phospho-

pyruvate carboxykinase, which converts oxalacetate to phosphopyruvate is located in the cytosol However, the mitochondrial membrane is relatively impermeable to oxalacetate so that this intermediate must be converted to aspartate, malate and/or fumarate which pass through the membrane and are reconverted to oxalacetate in the cytosol (Lardy, 1966) In the case of malate, this may also provide a reducing equivalent shuttle across the mitochondrial membrane The mitochondrial membrane is impermeable to reduced nicotinamide adenine dinucleotide (NADH) which is produced during fatty acid oxidation and which is used in the conversion of oxalacetate to malate In the cytosol malate is converted to oxalacetate yielding NADH which is then available to facilitate the conversion of 1,3-diphosphoglyceric acid to glyceraldehyde 3-phosphate (Fig 7 110)

Table 7 16
Gluconeogenesis in the postabsorptive state and after a prolonged fast

	Glucose Production		
	Rate	Source	
		Liver	Kidney
	g /day		
Postabsorptive	150–300	>90%	<10%
5 to 6 Week fast	86	55%	45%

From Felig et al , 1970

GENERAL FEATURES OF CONTROL

The role of gluconeogenesis in the maintenance of the blood glucose level requires that the activity of the pathway is sensitive to changes in the hormonal and dietary status of the animal Control of gluconeogenesis occurs at many levels but the nature of many of the control mechanisms is not well-understood Studies on the control of gluconeogenesis have dealt mainly with the process in liver and the material presented here will be limited to results obtained with this organ In general, control may occur in peripheral tissues to alter the supply of glucose precursors or effectors (hormones, fatty acids, etc) to the liver or may operate in the liver to control the uptake of precursors or their conversion to glucose Rapid alterations in the rate of gluconeogenesis may result from changes in the blood levels of substrates or hormones reaching the

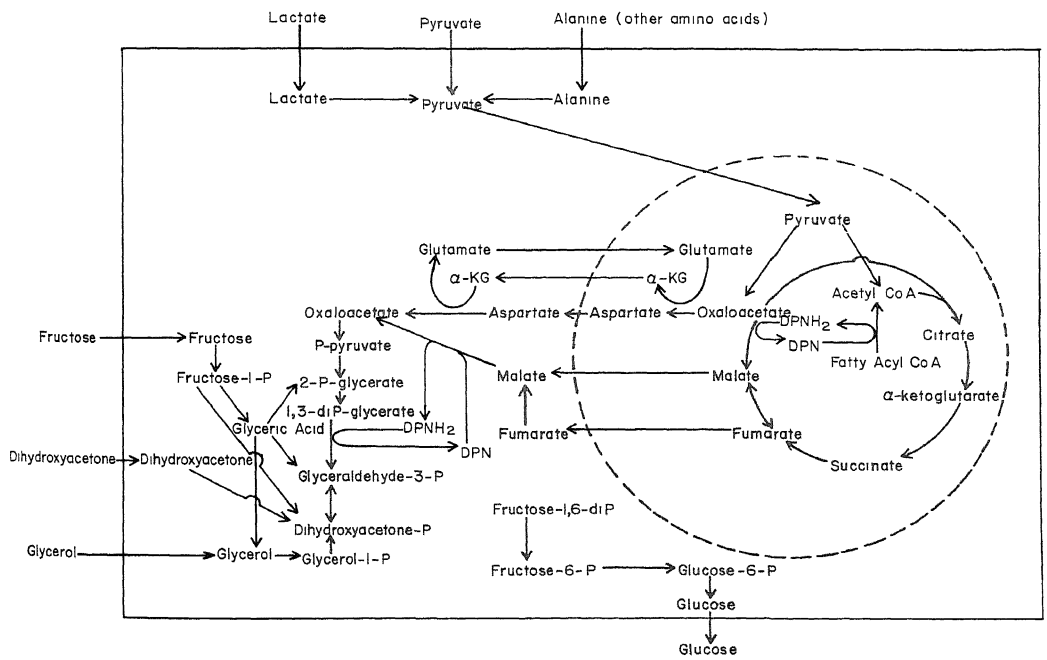


Fig 7 110 Diagram of the pathways of gluconeogenesis from lactate pyruvate alanine fructose and dihydroxy acetone in the liver A liver cell is represented by the rectangle and a mitochondrion by the circle

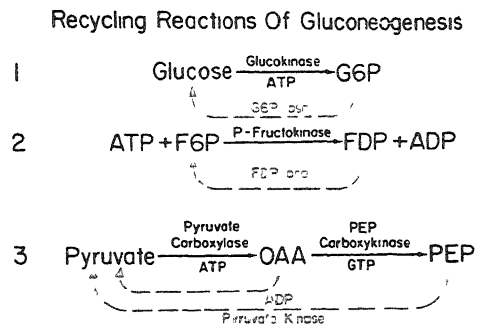


Fig 7 111 Thermodynamically irreversible steps of glycolysis and gluconeogenesis (From Williamson 1967)

liver or by allosteric-type regulation of the activity of gluconeogenic enzymes Slower alterations are observed which result from changes in the maximal gluconeogenic capacity of the tissue and which represent changes in the rate of enzyme synthesis (Weber, 1963) For example, stimulation of gluconeogenesis occurs during muscular exercise as a result of increased lactate production and during starvation as a result of increased availability of the gluconeogenic amino acids and glycerol However, during prolonged exposure to conditions requiring elevated rates of gluconeogenesis, increases in the maximal capacity of the tissue are observed The gluconeogenic flux is also regulated by the hormonal balance of the animal (e g , the rate is increased in diabetes)

Investigators have used whole animals, isolated perfused liver and kidney preparations, tissue slices, tissue homogenates, mitochondrial preparations, purified enzyme systems, and the measurement of adaptive enzyme changes in attempting to understand the control of gluconeogenesis (for review, see Newsholme and Gevers, 1967) Changes in gluconeogenic rates or capacities observed in the whole animal are difficult to interpret since both the gluconeogenic and the peripheral tissues may be involved Isolated perfused organ systems offer a number of advantages for studying the control of gluconeogenesis Hormones and substrates are delivered to the cells by the normal capillary bed, and therefore a direct effect on that tissue can be demonstrated since other tissues or organs are not included within the circulation In addition, samples of the tissue and perfusate can be obtained rapidly and with ease for the estimation of enzyme activities, metabolic intermediates, and substrate levels Rates of gluconeogenesis observed in perfused rat liver and kidney are in the same range as those estimated from studies of whole animals and suggest that the perfused preparations approximate the in vivo situation

CONTROL BY SUBSTRATE AVAILABILITY

Studies of Exton et al (1967, 1970) show that the gluconeogenic capacity of the liver is not saturated by the concentrations of gluconeogenic precursors normally present in the blood, indicating that the rate of glucose synthesis by the liver could be controlled by variations in plasma levels of lactate, pyruvate, amino acid, and glycerol (Fig 7 112 and 7 113) When the liver is presented with a mixture of physiological substrates at their normal blood concentrations, amino acids make the greatest contribution to glucose synthesis (Table 7 17) Any physiological change which would increase the blood lactate above its resting level of 1 mM should lead to increased glucose formation by the liver (Fig 7 113) Increased lactate production during exercise and during increased sympathetic nervous activity would be

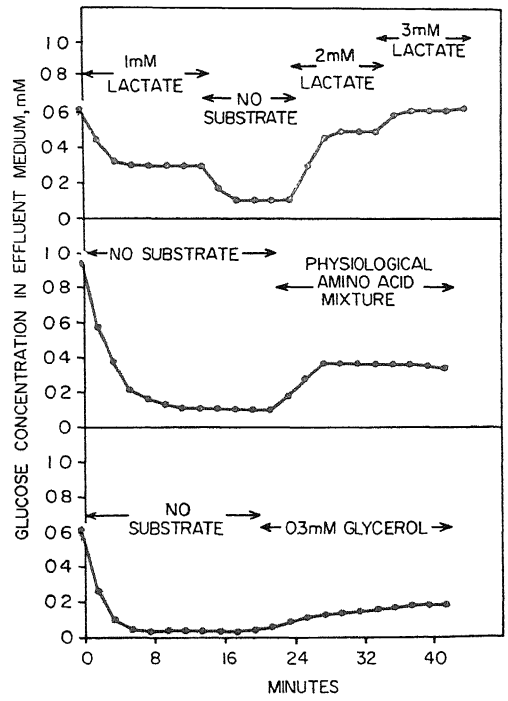


Fig 7 112 Effects of altering concentrations of substrates within the physiological range on gluconeogenesis Livers from fasted rats were perfused with nonrecirculating medium (From Exton et al 1970)

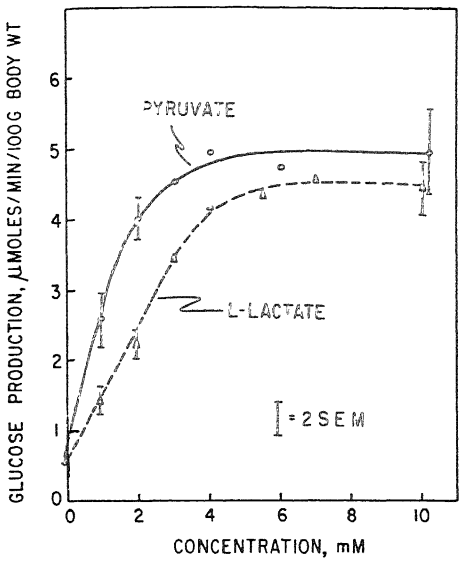


Fig 7 113 Relationship between gluconeogenesis and lactate or pyruvate concentration Livers from fasted rats were perfused with nonrecirculating medium containing lactate or pyruvate at the concentrations shown (From Exton and Park 1967)

accompanied by glucose synthesis, thus minimizing lactic acidosis and replenishing the glycogen stores of muscle and liver This process is known as the Cori cycle and, although it is physiologically important, it does not, strictly speaking, lead to formation of "new glucose" since lactate is itself derived from glucose (Cori, 1931) When the dietary supply of glucose does not satisfy the metabolic demands of the animal (e g , starvation, low carbohydrate diet), the gluconeogenic substrates utilized for the synthesis of new glucose are *1*) glycerol released from the triglyceride stores of adipose tissue, and *2*) amino acids derived from muscle protein In this respect protein is quantitatively the most important source of glucose for survival during fasting

The fact that hepatic gluconeogenesis can be markedly altered by variations in the levels of substrates within the physiological range illustrates the importance of regulation of amino acid and glycerol release from peripheral tissues in control of gluconeogenesis The prominent role of amino acids as substrates for gluconeogenesis suggests that the control of their release from muscle tissue would provide an important point of regulation Net amino acid release from peripheral tissues can result from hormonal control of either the rate of protein synthesis or the rate of protein degradation (chap 13) in that tissue Most studies have failed to differentiate between these effects

Insulin stimulates amino acid uptake and incorporation into protein in muscle and lowers the plasma α-amino nitrogen level The combination of a reduced availability of amino acids to the liver with a direct suppressing action of the hormone on hepatic gluconeogenesis (to be discussed in a subsequent section) renders insulin very inhibitory to amino acid gluconeogenesis The increased gluconeogenesis observed with diabetes is undoubtedly due partly to the mobilization of tissue protein which gives rise to increased plasma levels of most amino acids

Glucocorticoids exhibit a net protein catabolic effect in vivo They decrease amino acid uptake and incorporation into muscle protein and increase plasma α-amino nitrogen in eviscerated animals Glucocorticoids also stimulate utilization of amino acids by liver, but their peripheral effects appear to be stronger than their hepatic effects since they increase the plasma levels of most amino acids in adrenalectomized animals

Growth hormone increases amino acid uptake and incorporation into protein in muscle and decreases plasma α-amino nitrogen Inhibition of amino acid gluconeogenesis would therefore be expected, but the hormone has been observed to increase hepatic glucose output in intact dogs and in the perfused rat liver Further studies are necessary to define

Table 7 17
Rates of gluconeogenesis from substrates at concentrations found in rat blood plasma

Substrate	Rate of Glucose Production
	μmoles/ 100 g /hr
Lactate, 1 0 mM	51
Pyruvate, 0 1 mM	12
Amino acids[a]	91
Glycerol, 0 3 mM	18
Sum of above	172
Mixture of above substrates	170

From Exton and Park, 1967
[a] Amino acids were present at the concentrations found in rat blood plasma

the effect of growth hormone on amino acid gluconeogenesis

Effects of epinephrine and norepinephrine on peripheral amino acid release are uncertain

In fasting, release of amino acids from peripheral tissues is important for hepatic gluconeogenesis The situation seems quite complex, at least in man (Marliss et al , 1970) Early in starvation, the plasma levels of many amino acids increase and gluconeogenesis is activated This indicates increased protein catabolism probably due to decreased insulin secretion When starvation is prolonged, protein is conserved, the plasma levels of most amino acids decline, and gluconeogenesis is markedly attenuated The rate-limiting process at this time appears to be the release of substrate from peripheral tissues since exogenous amino acid is rapidly converted to glucose, suggesting that hepatic gluconeogenic mechanisms remain intact The decreased rate of gluconeogenesis during prolonged fasting is accompanied by the adaptation of the brain to utilize ketone bodies (acetoacetate and β-hydroxybutyrate) in addition to glucose (Owen et al , 1967)

Studies of the individual amino acids as gluconeogenic substrates indicate that alanine is the major precursor of glucose in the liver with lesser contributions from serine, proline, threonine, glutamate, glutamine, and asparagine (Ross et al , 1967) Other studies indicate that plasma alanine concentrations undergo the greatest alterations when release of amino acids from peripheral tissue is affected (Felig et al , 1970) These findings have led to the proposal of the existence of an "alanine cycle" which could be of importance in nitrogen metabolism As depicted in Figure 7 114, the transfer of amino groups from several amino acids to pyruvate results in the formation of alanine in extrahepatic tissues The donor amino acids come from blood or tissue protein and pyruvate is formed from blood glucose or tissue glycogen Newly formed alanine and alanine derived from protein breakdown is released and carried to the liver where it is used for the synthesis of urea (which is excreted) and glucose (which returns to the peripheral tissue) This process provides a means by which amino acids can be rapidly catabolized in extrahepatic tissues without the release of toxic amounts of ammonia

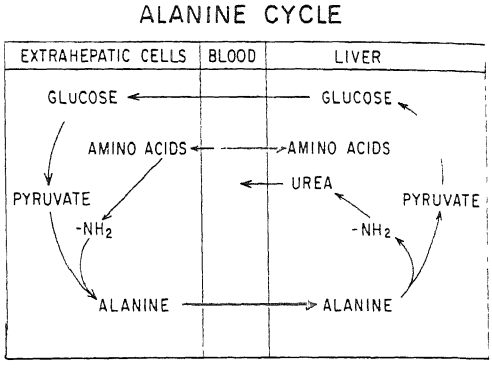


Fig 7 114 Diagrammatic representation of the postulated role of alanine in the transfer of amino groups from extrahepatic tissues to the liver (From Mallette et al 1969)

HORMONAL CONTROL OF GLUCONEOGENESIS

Control of gluconeogenesis by the direct action of hormones on the liver has been studied intensively in the isolated perfused rat liver It has been possible to demonstrate that several hormones act directly at the hepatic level and that some of these (glucagon, epinephrine, insulin) act within seconds or minutes while others act more slowly (glucocorticoids) The slower effects may involve changes in enzyme synthesis while the fast effects involve regulation of the intracellular level of adenosine 3′,5′-monophosphate (cyclic AMP) This compound has been implicated in the regulation of a number of cellular processes (chap 1)

Glucagon plays a major role in the regulation of blood glucose levels through its effects on hepatic glycogenolysis and gluconeogenesis In fact, it probably has no action of physiological importance outside the liver and pancreas The effect of the hormone on glycogenolysis was recognized early (for review, see Sutherland and Rall, 1960), and recent studies have confirmed that glucagon is also an important regulator of the gluconeogenic pathway The hormone increases the conversion of lactate, pyruvate, alanine, and bicarbonate to glucose in the perfused liver The effect is apparent within 40 sec after addition to the perfusate and concentrations of the hormone within the physiological range are effective (Fig 7 115) The effects of glucagon are thought to result from activation of membrane bound adenyl cyclase with the ensuing intracellular accumu-

lation of cyclic AMP The following lines of evidence indicate that the effect of glucagon on gluconeogenesis is mediated by increased cyclic AMP levels *1*) the hormone is capable of stimulating gluconeogenesis only in concentrations which increase tissue level of cyclic AMP, 2) the increase in cyclic AMP precedes the activation of gluconeogenesis, *3*) direct addition of cyclic AMP to the perfusate mimicks the effects of glucagon, and *4*) exogenous cyclic AMP produces the same changes in metabolite levels in the gluconeogenic pathway as does glucagon (i e , a "crossover point" is observed between pyruvate and P-pyruvate, the site of the rate-limiting reactions from lactate, pyruvate, or alanine gluconeogenesis) (Fig 7 116) The stimulatory effects of glucagon and cyclic AMP on the conversion of pyruvate to phosphopyruvate appear to be too rapid to involve changes in enzyme synthesis As shown in Figure 7 110, the conversion of pyruvate to phosphopyruvate is thought to involve a mitochondrial enzyme, pyruvate carboxylase, and a cytosol enzyme, phosphopyruvate carboxykinase A stimulation could involve either or both of these enzymes or alternatively could represent a change in the rate of flux of intermediates across the mitochondrial membrane A direct effect of glucagon or cyclic AMP on the activity of either enzyme has not been demonstrated In addition to these effects, glucagon and cyclic AMP stimulate the entry of alanine into the liver cell There is some evidence for an effect of cyclic AMP on the transfer of substrates across mitochondrial membranes, but the importance of this effect has not been examined in detail

Numerous studies suggest that increased

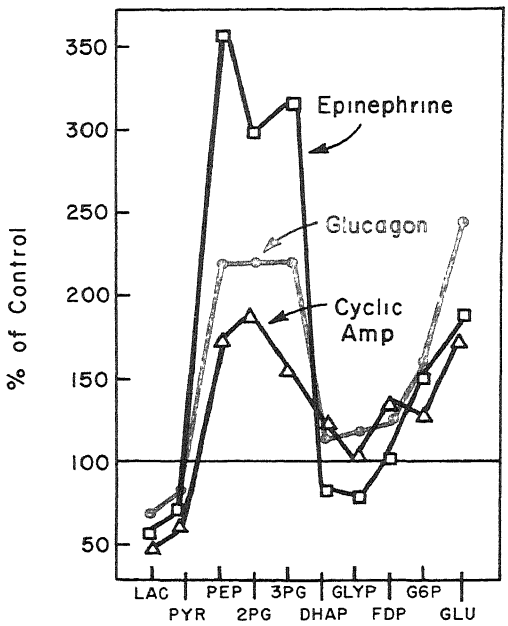


Fig 7 116 Effects of glucagon epinephrine and cyclic AMP on levels of gluconeogenic intermediates in perfused livers from fasted rats Values are expressed as percentages of those found in livers perfused with lactate alone Concentrations of additions were glucagon 1 × 10^{-8}M epinephrine 1 × 10^{-6}M cyclic AMP 1 × 10^{-3}M The intermediates listed along the abscissa are lactate (LAC) pyruvate (PYR) P-pyruvate (PEP) 2-P-glycerate (2PG) 3-P-glycerate (3PG) dihydroxyacetone-P (DHAP) glycerol-1-P (GLYP) fructose-1 6-di-P (FDP) glucose-6-P (G6P) and glucose (GLU) This is a method employed to identify the specific step in the gluconeogenic pathway which is stimulated by glucagon epinephrine and cyclic AMP Measurement of changes in the concentration of intermediates along the pathway when the rate of flux is altered should indicate the rate-limiting reactions Stimulation of the pathway may lead to depletion of substrate proximal to the site of regulation and accumulation of substrate distal to this site The step at which the change from depletion to accumulation occurs is called the crossover point and identifies the rate-controlling reaction The data presented in this figure indicate that glucagon epinephrine and cyclic AMP all enhance the rate of gluconeogenesis by stimulating the conversion of pyruvate to P-pyruvate (Drawn from data presented by Exton and Park 1969)

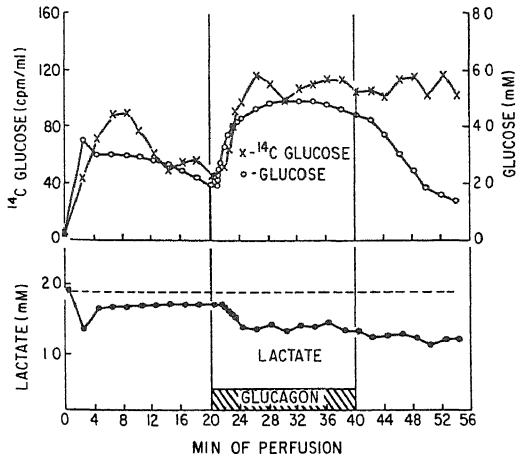


Fig 7 115 Effects of glucagon on glucose output lactate uptake and glucose-^{14}C synthesis in livers from fed rats perfused with nonrecirculating medium containing lactate-^{14}C Glucagon was infused from 20 to 40 min to produce a concentration 1 × 10^{-8}M in the influent medium (From Exton and Park 1968)

availability of fatty acids coupled to enhanced fatty acid oxidation stimulates hepatic gluconeogenesis (Williamson et al, 1966, 1969a, 1969b) Stimulation of two sites in the gluconeogenic pathway has been implicated *1)* mitochondrial enzyme, pyruvate carboxylase which is absolutely dependent upon acetyl CoA for activity, and *2)* the triose phosphate step, glyceraldehyde-P dehydrogenase, which requires reduced nicotinamide adenine dinucleotide (NADH) (Fig 7 110) Increased amounts of both acetyl CoA and NADH are generated during accelerated β-oxidation of fatty acids This has led to the suggestion that the effect of glucagon on gluconeogenesis is the result of increased fatty acid oxidation due to activation of a hepatic lipase enzyme (Struck et al, 1965) According to this proposal, glucagon would increase the intracellular level of cyclic AMP which would activate an hepatic lipase Increased lipase activity would accelerate hydrolysis of endogenous triglyceride and enhance β-oxidation of fatty acids and result in generation of both acetyl CoA and NADH The following lines of evidence support this proposal *1)* glucagon pretreatment reduces hepatic fat content and increases lipase activity of crude liver homogenates, *2)* addition of fatty acids to isolated liver preparations can stimulate gluconeogenesis under some conditions, and *3)* acetyl CoA levels correlate with the rate of gluconeogenesis

Although this proposal is held in high esteem by a number of investigators, others question it as an adequate explanation of the gluconeogenic action of glucagon The following findings do not fit with the proposal *1)* glucagon frequently has no effects on ketogenesis even though it markedly stimulates glycogenolysis and gluconeogenesis, *2)* fatty acids are never as effective as glucagon in stimulating gluconeogenesis in the perfused liver even though they increase ketogenesis to a much greater degree than does glucagon, *3)* the gluconeogenic effect of fatty acids is additive with that of glucagon, and *4)* fatty acids produce different changes in the levels of gluconeogenic intermediates than does glucagon It seems likely that further research will determine which mechanism is of physiological importance

An additional possibility for the site of action of glucagon is the pyruvate dehydrogenase complex Recent evidence suggests that this enzyme exists in an active dephosphorylated form and that its activity disappears upon phosphorylation (Linn et al, 1969) According to this scheme, cyclic AMP would activate the protein kinase responsible for phosphorylating the active form of the enzyme to render it inactive Thus, cyclic AMP would diminish carbon flow through pyruvate dehydrogenase and direct it toward glucose synthesis However, attempts to demonstrate such an effect of cyclic AMP have not been successful to date

Effects of epinephrine and norepinephrine on glycogenolysis and gluconeogenesis are similar to those produced by glucagon and are thought to involve the same mechanisms of action (i e, via cyclic AMP) The concentrations of epinephrine and norepinephrine required to elicit maximal effects in the perfused liver are considerably higher than the observed plasma levels of these hormones, raising the possibility that hepatic glucose production may not be controlled by circulating levels of catecholamines as has been suggested However, this does not negate the important role of the sympathetic nervous system in the regulation of glucose output by the liver in which catecholamine is released from sympathetic nerve endings

Insulin inhibits gluconeogenesis As mentioned earlier, it diminishes the supply of gluconeogenic substrates from tissues such as muscle and fat and, in addition, the hormone directly inhibits glucose synthesis in the liver Insulin deficiency (diabetes) and short-term starvation (a relative insulin lack) cause increased gluconeogenesis The mechanisms involved in these effects were suggested from studies using animals made acutely diabetic by injection of anti-insulin serum (Jefferson et al, 1968) Perfused livers from rats treated with anti-insulin serum showed increased rates of glycogenolysis and gluconeogenesis and these effects were abolished by insulin addition in vitro The similarity of the effects produced by anti-insulin serum to those caused by exogenous cyclic AMP led to an investigation as to whether these effects resulted from an increase in the tissue level of cyclic AMP It was found that anti-insulin serum caused a progressive rise in liver cyclic AMP over 1 hr of treatment which persisted during a subse-

quent perfusion Insulin added to the perfusate reduced the elevation in cyclic AMP caused by anti-insulin serum treatment and produced a small reduction in the level of the nucleotide in livers from untreated animals It was also found that liver cyclic AMP levels in vivo were doubled when diabetes was induced by alloxan and were increased by 60% by starvation but could be rapidly lowered to normal by insulin injection These studies suggest that insulin can exert direct inhibitory effects on gluconeogenesis and glycogenolysis by lowering the level of cyclic AMP Other hepatic actions of insulin may also be related to the reduction in cyclic AMP but this is uncertain at this time In addition, these studies suggest that the stimulation of gluconeogenesis during diabetes and short-term starvation can be ascribed partly to the increase in cyclic AMP However, changes in the levels of enzymes also play an important role in the long-term adaptations of diabetes and starvation

A role for the glucocorticoids in the regulation of hepatic glucose synthesis was suggested by early studies which showed that the fatal hypoglycemia of fasted adrenalectomized animals was due to a failure of gluconeogenesis The mechanism of action of adrenal steroids on gluconeogenesis is unclear, however As discussed earlier, glucocorticoids influence the flow of glucogenic substrates to the liver In addition they appear to have direct effects on liver The direct effects of adrenal steroids on the liver may involve changes in enzyme synthesis Recent studies have shown that adrenal deficiency has little effect on gluconeogenesis in fed animals but causes the pathway to be unresponsive to the elevated level of cyclic AMP associated with starvation and diabetes This has led to the suggestion that adrenal steroids exert mainly a "permissive action" on gluconeogenesis so that their effects do not become evident unless the tissue level of cyclic AMP is increased as a result of another hormone action (e g , glucagon) The following evidence supports this suggestion *1*) glucagon is ineffective in stimulating glucose synthesis in livers from fasted adrenalectomized rats, *2*) treatment with glucocorticoids in vivo or in vitro restores to normal the gluconeogenic response to glucagon, and *3*) the activation of adenyl cyclase by glucagon is not impaired in the absence of glucocorticoids

Growth hormone, thyroxine, and other hormones may play a role in the regulation of gluconeogenesis but their mechanisms of action are not known at this time

Ketogenesis

Ketogenesis is the term applied to the metabolic process involving the production of the ketone bodies, acetoacetate, acetone, and β-hydroxybutyrate It is a normal function of the liver to produce these metabolites for peripheral utilization Whenever ketogenesis by the liver exceeds ketone body utilization by other tissues, these compounds accumulate in the extracellular fluid leading to a state termed ketosis This situation has important clinical significance since high levels of ketone bodies cause acidosis and exert toxic effects on tissues such as the brain On the other hand, the ability of the liver to produce ketone bodies instead of carbon dioxide as an end product of metabolism has important physiological significance in coordinating the energetic needs of the organism as a whole Acetoacetate and β-hydroxybutyrate are used by muscle and perhaps other tissues as a source of energy in preference to other substrates Therefore, from a physiological point of view, the liver is converting the energy available in long-chain fatty acids to a more readily utilizable nutrient for other tissues

The metabolic pathway for ketogenesis in the liver is shown in Figure 7 117 It can be seen that ketogenesis is inseparable from the β-oxidation of fatty acids and that the enzymatic machinery is localized mainly within the mitochondria (Bressler, 1963, Wieland, 1968) Acetyl CoA, an intermediate derived mainly from the β-oxidation of long-chain fatty acids, is the immediate precursor of acetoacetate Formation of the ketone bodies is initiated by the condensation of two molecules of acetyl CoA to form acetoacetyl CoA This step is catalyzed by the enzyme β-ketoacylthiolase. The next step in the enzymatic formation of acetoacetate involves the condensation of another molecule of acetyl CoA with acetoacetyl CoA to give β-hydroxy-β-methyl-glutaryl-CoA (HMG-CoA) Enzymatic cleavage of this molecule yields acetoacetate and acetyl CoA This process is called the HMG-CoA cycle

The other two ketone bodies are derived

PATHWAY OF KETOGENESIS

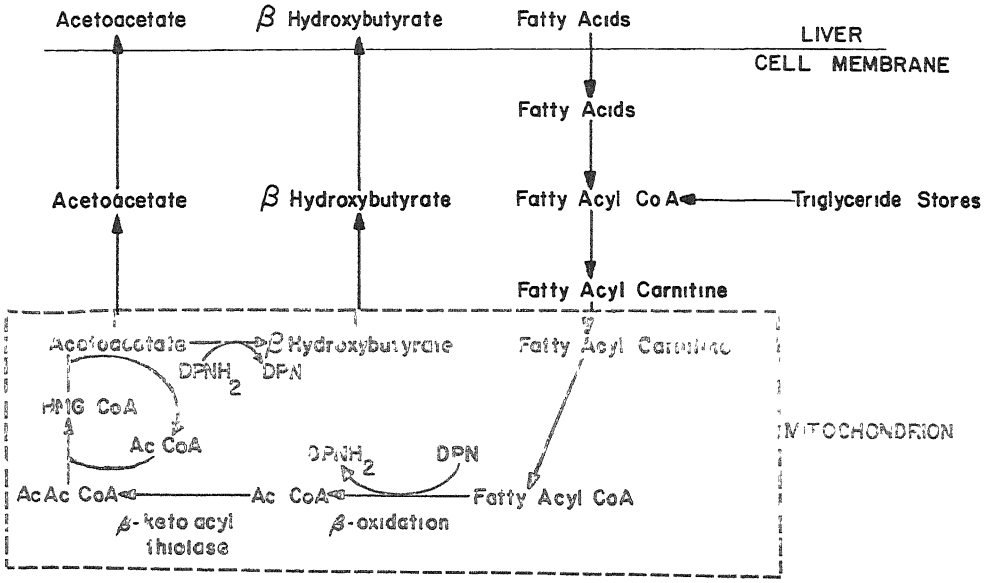


Fig 7 117 Diagram of the pathways of ketogenesis from fatty acids in the liver

from acetoacetate Acetone apparently originates by spontaneous decarboxylation of acetoacetate Reduction of acetoacetate in the NADH-dependent reaction catalyzed by the enzyme β-hydroxybutyrate dehydrogenase gives rise to β-hydroxybutyrate This compound comprises 65 to 75% of the total ketone bodies present in blood

Before acetoacetate can enter oxidative metabolism in peripheral tissues, it must be reactivated to acetoacetyl CoA by an enzymatic conversion involving succinyl CoA The liver is relatively incapable of carrying out this reaction For peripheral utilization, β-hydroxybutyrate is oxidized to acetoacetate which then undergoes the reactions described above

In general, the production of ketone bodies is stimulated whenever the availability of carbohydrate is relatively (e g , diabetes) or absolutely (e g., fasting) deficient The degree of ketosis that develops depends on the disproportion between fat catabolism and carbohydrate utilization A number of circumstances which may contribute to the development of ketosis are listed in Table 7.18 It is clear that the rate of ketone body production is a function of *1*) the rate of mobilization of free fatty acids from adipose tissue, *2*) the rate of uptake and accumulation of lipid by the liver, *3*) the rate of fatty acid oxidation in relation to carbohydrate availability in the liver, and *4*) the rate of carbohydrate utilization by all the tissues of the body In addition, the blood concentration of ketone bodies is a function of the rate of their utilization by muscle and other extrahepatic tissues Therefore, ketogenesis and the development of ketosis may be affected by regulation at a number of levels Since the liver takes up triglycerides and fatty acids in proportion to their concentrations in the blood, any of the factors which control the release of fatty acids from adipose tissue would be involved in the regulation of ketogenesis

Lipogenesis

During feeding there is need to store much of food energy for use in the postabsorptive state by converting it to fat, an energy store form Lipogenesis is the process concerned with conversion of glucose and metabolites such as pyruvate and acetyl CoA to triglyceride (Fig 7 118) Since gluconeogenesis and lipogenesis share all of the reversible enzymes between phosphopyruvate (PEP) and glucose-6-P, it is apparent that when one process is active the other must be suppressed

In the fed state, blood levels of glucose and insulin are high and there is decreased availability of fatty acids to the liver These conditions favor restraint of gluconeogenesis and provide more substrate for lipogenesis by increasing intracellular glucose concentrations in the liver An increased phosphorylation of glucose by glucokinase yields glucose-6-P which is converted to pyruvate Pyruvate enters the mitochondria and undergoes oxidation to acetyl CoA, the major substrate for fat synthesis Since acetyl CoA does not diffuse readily into the cytosol, where fatty acid synthesis occurs, it must be converted to citrate by condensation with oxaloacetate within the mitochondria Conversion of pyruvate to acetyl CoA, rather than to oxalocetate, is favored by the low intramitochondrial acetyl CoA levels that exist in the fed state Decreased availability of fatty acids to the liver results in a fall in intracellular fatty acyl CoA concentration which in turn leads to decreased intramitochondrial oxidation of fatty acids and a fall in intramitochondrial acetyl CoA levels This favors a decrease in the activity of pyruvate carboxylase, the enzyme responsible for catalyzing formation of oxaloacetate, and a release of inhibition of pyruvate dehydrogenase, which allows for more rapid formation of acetyl CoA and subsequently citrate Citrate is transported into the cytosol where it is cleaved to acetyl CoA and oxaloacetate by the citrate cleavage enzyme This enzyme, in addition to providing acetyl CoA for cytoplasmic fat syn-

Table 7 18
Causes of ketosis[a]

1) Carbohydrate deficiency or loss, hypoglycemia
   a) High fat, low carbohydrate diet
   b) Starvation
   c) Insulin hypoglycemia
   d) Renal glycosuria
   e) Glycogen storage disease
2) Increased metabolic requirements
   a) Pregnancy
   b) Lactation
   c) Rapid growth
   d) Convalescence and wound healing
   e) Vigorous exercise
   f) Cold exposure
   g) Fever
   h) Hyperthyroidism
3) Metabolic blocks
   a) Diabetes mellitus
   b) Sodium fluoroacetate, malonate, NH_4Cl

From Engel and Amatruda, 1963

[a] A list of some of the circumstances in which lipid catabolism is disproportionately great in relation to carbohydrate utilization and in which, therefore, ketosis may be expected to occur

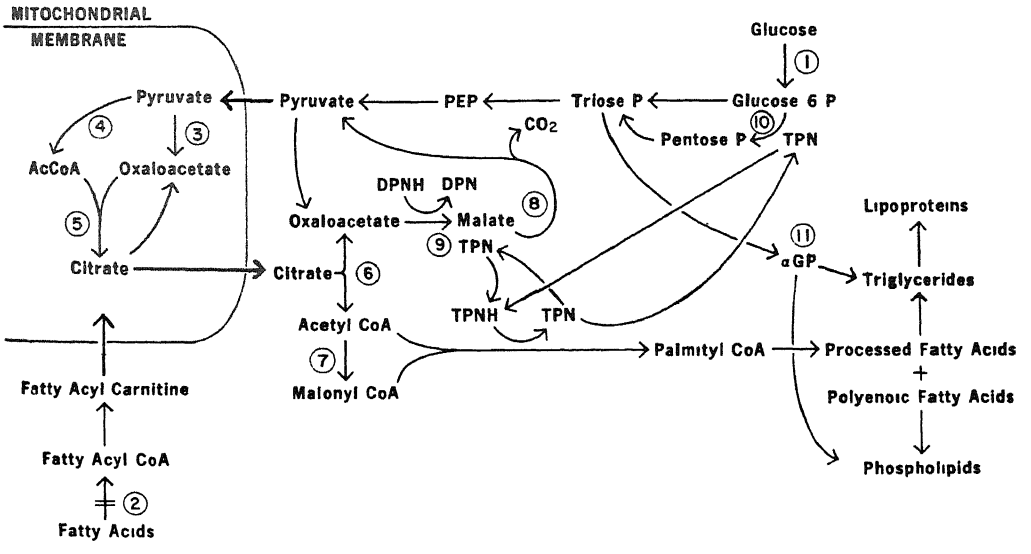


Fig 7 118 Control of the lipogenic pathway Conditions favoring lipogenesis *1)* increased glucokinase activity *2)* decreased availability of fatty acids *3)* pyruvate carboxylase activity decreased, *4)* pyruvate dehydrogenase inhibition released *5)* citrate synthetase active *6)* citrate cleavage enzyme *7)* acetyl-CoA carboxylase inhibition released *8)* malic enzyme active *9)* initiation of transhydrogenase sequence yielding TPNH *10)* pentose phosphate shunt pathway active yielding TPNH *11)* rapid generation of α-glycerophosphate (From Tepperman and Tepperman 1970)

thesis, functions together with the NADP-requiring malic enzyme to retrieve pyruvate for recycling and to produce NADPH which is required to supply reducing equivalents for lipogenesis. Additional NADPH is generated through the pentose shunt.

The rate-limiting reaction in the lipogenic pathway is catalyzed by acetyl CoA carboxylase. This enzyme is competitively inhibited by high intracellular concentrations of fatty acyl CoA esters. Since these decrease in the fed state, inhibition of acetyl CoA carboxylase is released and fatty acid synthesis is favored. Inhibiting concentrations of fatty acyl CoA do not accumulate during active fatty acid synthesis because the rapid rate of generation of α-glycerophosphate favors sequestration of newly synthesized fatty acids into triglyceride.

The rate of lipogenesis is controlled by the nutritional state of the animal. The rate is high in the fed state and is depressed during restricted food intake, by a high fat diet, or by a deficiency of insulin, as in diabetes. The steps that are numbered in Figure 7.118 indicate acute changes which favor an increased rate of lipogenesis. In fasting, lipogenesis is inhibited by *1*) decreased availability of glucose and decreased glucokinase levels; *2*) decreased availability of α-glycerophosphate; and *3*) increased levels of fatty acyl CoA which inhibits acetyl CoA carboxylase and provides for higher intramitochondrial levels of acetyl CoA. In addition to these acute mechanisms, the rate of lipogenesis can be affected by changes in synthesis of other enzymes in the pathway.

13 / HORMONAL CONTROL OF GROWTH AND PROTEIN METABOLISM

Hormonal control of protein and nucleic acid metabolism is expressed in the growth of single cells, organs, and the body Effects of hypophysectomy, thyroidectomy, and castration to retard growth indicate that hormonal control is extensive With the exception of growth hormone, structure, synthesis, release, and general metabolic effects of these hormones have been considered in earlier chapters In this chapter, control of growth and protein metabolism will be discussed, and general features of the action of growth hormone will be described

GROWTH HORMONE

Growth hormone is synthesized and stored in acidophilic cells of the anterior pituitary Secretion granules containing the hormone have been isolated Release of hormone by these cells is controlled by growth hormone-releasing factor (GRF) (chap 2) Human growth hormone is made up of 188 amino acids and has a molecular weight of 21,500 (Fig 7 119) Development of an immunoassay for growth hormone has permitted studies of various factors that control growth hormone release by GRF These factors fall into three groups *1*) Decreased energy supply, as encountered in fasting, exercise and hypoglycemia, increases plasma growth hormone levels in direct relation to the fall in blood sugar (Fig 7 120) Release of growth hormone in situations of reduced energy supply results in release of fatty acid from adipose tissue and reduction in glucose utilization by muscle through restraint of membrane transport and intracellular glucose phosphorylation Patients that lack pituitary function have a prolonged period of hypoglycemia following insulin administration (Fig 7 121) Treatment with cortisone and thyroxine partially restores this defect, but growth hormone is required for a rapid return of blood sugar toward normal 2) Increased levels of plasma amino acids resulting from protein feeding or intravenous injection of amino acids increase plasma levels of growth hormone Arginine is particularly potent in its ability to release growth hormone, and, as shown in Figure 7 122, infusion of arginine increases plasma growth hormone levels to the same extent as insulin hypoglycemia Release of hormone in response to higher plasma levels of amino acids favor increased synthesis of protein and returns plasma amino acid levels to normal by facilitating entry of these compounds into cells *3*) Stress, such as surgical procedures, increases growth hormone release (Baylis et al , 1968)

Growth hormone is required for normal growth during adolescence Small stature may result from many factors, however, including nutritional, genetic, and other types of hormonal unbalance Deficiency of growth hormone results in dwarfism, while overproduction causes gigantism Gigantism due to growth hormone is characterized by uniform growth of all body components The extent of growth hormone deficiency in children of short stature can be estimated by stimulation of growth hormone secretion with infusions of arginine or insulin Peak levels of growth hormone that are achieved are an index of the adequacy of secretion (Parker and Daughaday, 1968) Treatment of hormone-deficient patients with human growth hormone increases growth from 1 to 2 cm /year to 7 to 12 cm /year Since effects of growth hormone are species specific, bovine or porcine hormone have little or no effects on growth in man

Excessive growth hormone secretion in adult life results in a condition known as *acromegaly* This condition is characterized by *1*) increased plasma levels of growth hormone (from 0 to 4 to 10 to 600 mμg /ml), 2) overgrowth of the bones of hands, feet, and face; and *3*) overgrowth of viscera, hair, and soft tissues of the body An adenoma of the anterior lobe of the

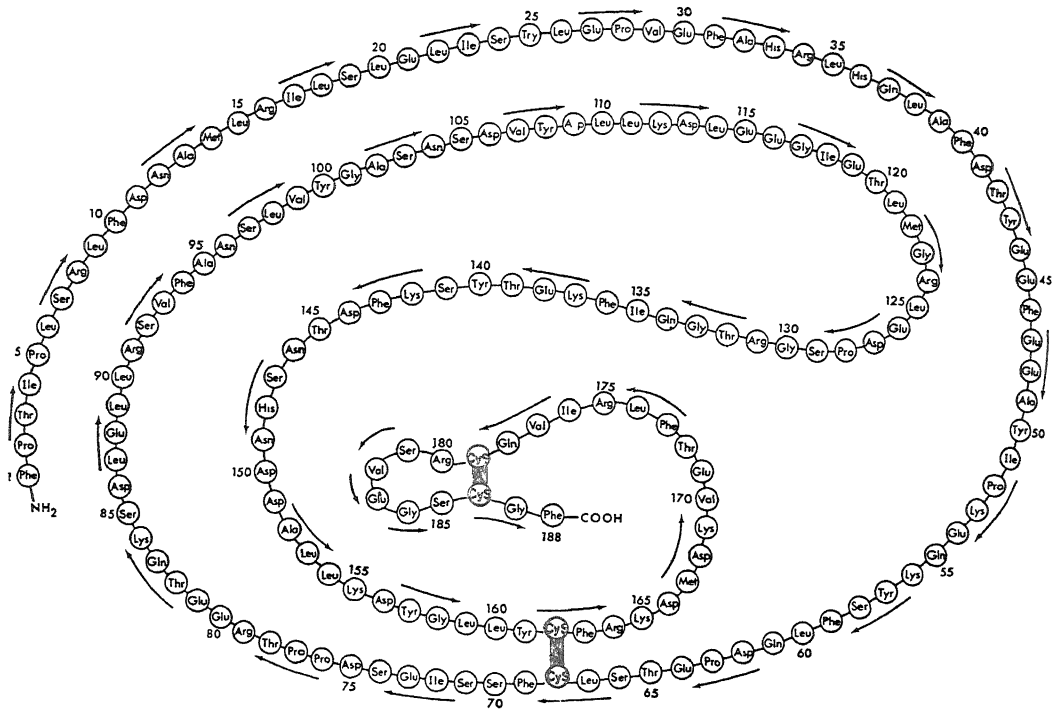


Fig 7 119 Amino acid sequence of human growth hormone Human growth hormone is a single polypeptide chain with two disulfide bridges When the disulfide bridges are reduced with mercaptoethanol in 8M urea and alkylated with iodoacetamide no loss of growth-promoting activity is found After partial digestion of the molecule with either pepsin or chymotrypsin biological activity in unchanged These findings suggest that it may be possible to synthesize a smaller peptide with growth-promoting activity (From Li 1968)

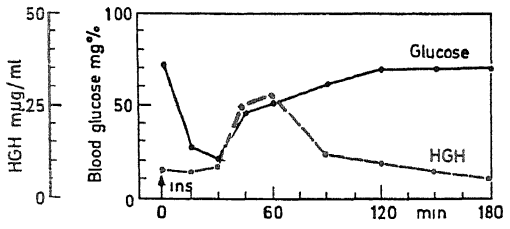


Fig 7 120 Effect of insulin hypoglycemia (0 05 units/kg body weight) on HGH concentration in plasma in a healthy subject (From Luft and Cerasi 1968)

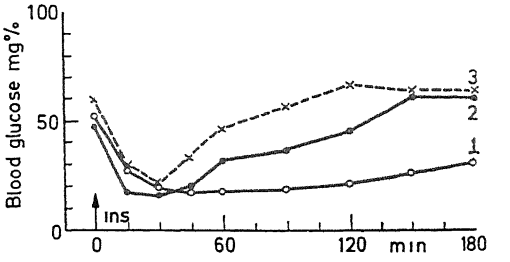


Fig 7 121 Insulin tolerance test (0 05 units/kg body weight) in a patient with panhypopituitarism *1)* without replacement *2)* on cortisone + thyroxine *3)* on cortisone + thyroxine + growth hormone (From Luft and Cerasi 1968)

pituitary is usually responsible for the hypersecretion of hormone Increased length of the long bones does not occur since the epiphyseal plates are already fused. Glycosuria and hyperglycemia are common and a condition indistinguishable from diabetes of pancreatic origin may be present. The acromegalic patients that develop hyperglycemia have a hereditary predisposition to diabetes and are unable to respond to the increased demands for insulin with an increased rate of secretion

Biological actions of growth hormone include a wide range of effects that cannot be explained by a single action of the hormone Addition of growth hormone to isolated muscles or adipose tissue of hypophysectomized rats accelerates entry of both amino acids and sugars into the cell (Fig 7 123) These effects are much less marked when muscles from normal animals are used These "insulin-like" effects of growth hormone may have little or

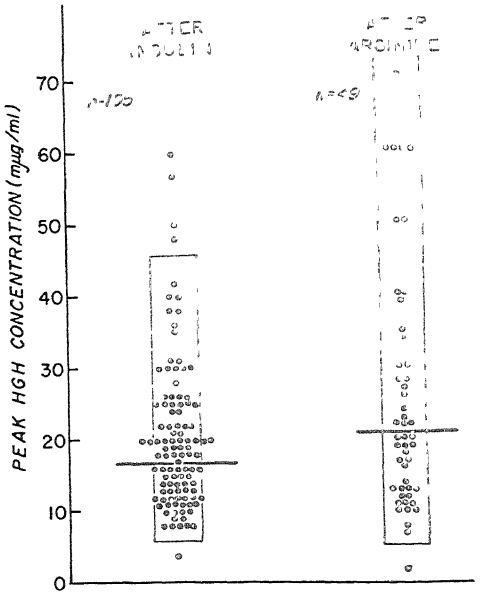


Fig 7 122 Comparison of peak growth hormone concentration of serum following insulin induced hypoglycemia and arginine infusion (From Parker et al 1967)

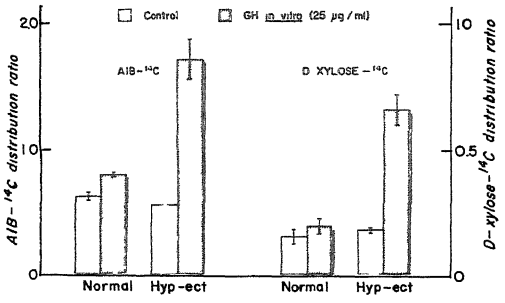


Fig 7 123 Effects of growth hormone in vitro on the uptake of α-aminoisobutyric acid (AIB-^{14}C) and of D-xylose-^{14}C by isolated diaphragms from normal and hypophysectomized rats (From Ahren and Hjalmarson 1968)

no physiological significance Administration of growth hormone to hypophysectomized rats for longer times decreases the sensitivity of glucose transport to stimulation by insulin (Fig 7 124) This effect is potentiated by the simultaneous administration of hydrocortisone An additional factor responsible for the diabetogenic effect of growth hormone is a block in glucose phosphorylation Even if sufficient insulin is administered to overcome the insensitivity of transport to insulin, glucose utilization is restrained by the block in phosphorylation As seen in Fig 7 125, phosphorylation activity is normal in the heart of hypophysectomized diabetic rats but is impaired in the heart of intact diabetic rats A severe block in

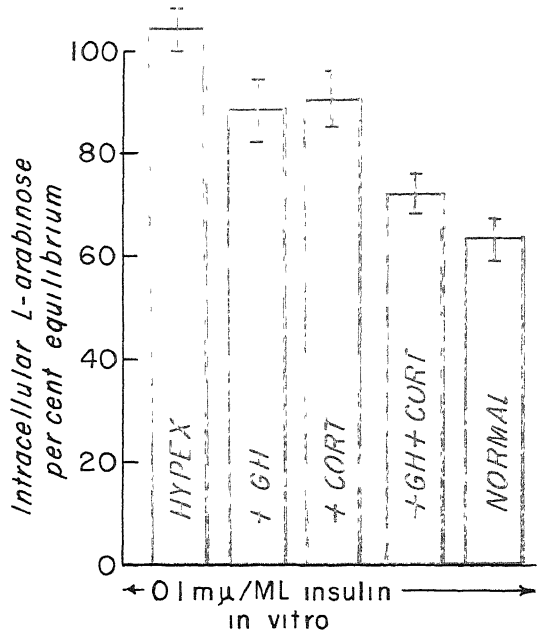


Fig 7 124 The late effect of growth hormone (GH) and hydrocortisone (Cort) in vivo on the sensitivity of transport to insulin as measured by L-arabinose accumulation Growth hormone was injected into hypophysectomized rats in a dose of 0 1 mg/100 g and cortisone in a dose of 2 5 mg/100 g 24 hr before the heart was removed for testing (From Park et al 1961)

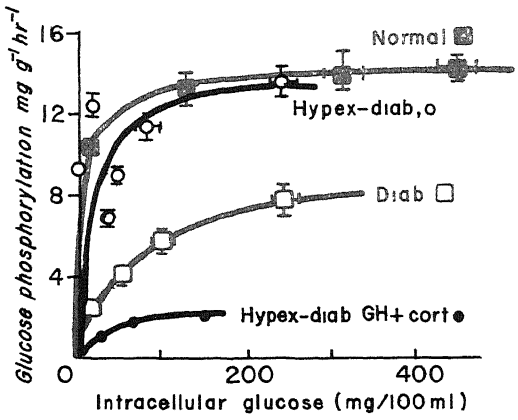


Fig 7 125 Effects of alloxan diabetes hypophysectomy growth hormone and hydrocortisone on the kinetics of glucose phosphorylation in the perfused heart Where indicated growth hormone (GH) and hydrocortisone (Cort) were injected over a 24-hr period before removal of the hearts for testing (From Morgan et al 1961)

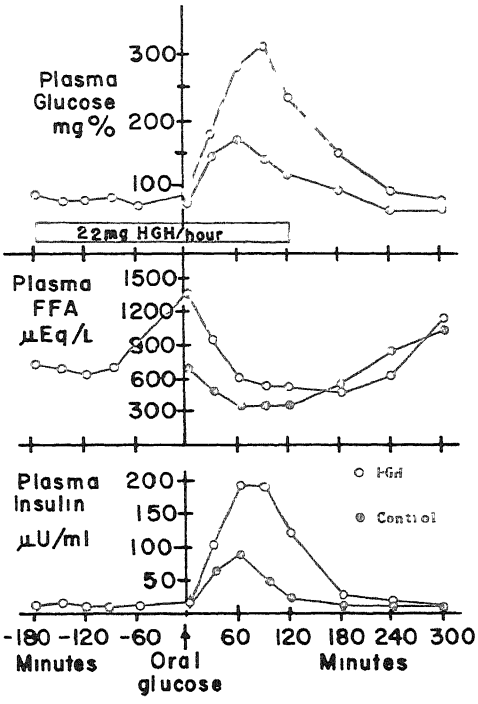


Fig 7 126 The effect of human growth hormone on glucose tolerance plasma free fatty acid (FFA) and insulin in a human subject (From Kipnis 1965)

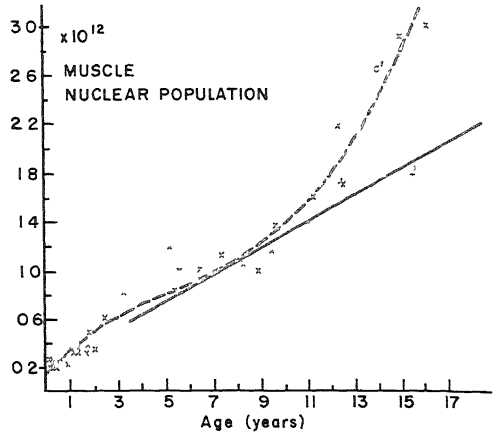


Fig 7 127 Relationship between nuclear number in the musculature is shown with chronological age for boys and girls Clearly the male adolescent growth spurt is more remarkable and associated with increase of nuclear number (From Cheek and Hill 1970)

phosphorylation develops when hypophysectomized diabetic rats are treated with growth hormone and hydrocortisone Since fatty acids reduce insulin sensitivity of glucose transport and block glucose phosphorylation (chap 10) impairment of these steps in growth hormone-treated animals may be related to the lipid-mobilizing effects of growth hormone Administration of human growth hormone to normal men for 3 hr increases plasma levels of free fatty acids (Fig 7 126) When a glucose tolerance test follows, plasma glucose increases to higher levels in hormone-treated subjects than in controls, despite higher plasma levels of insulin The return of plasma glucose to basal levels is slower after hormone treatment, which indicates a reduced sensitivity to insulin Whether growth hormone also has direct effects on muscle to control the rate of glucose transport and phosphorylation is undecided

Effects of growth hormone to accelerate growth appear to involve stimulated rates of ribonucleic acid (RNA) and protein synthesis These effects will be discussed along with the effects of other hormones on these pathways.

CONTROL OF CELLULAR GROWTH

Growth depends upon either an increase in the number (hyperplasia) or size (hypertrophy) of body cells Some tissues, such as muscle, loose the capacity for cell division while other tissues, such as liver, retain this capacity in adult life Since the size of any cell is limited, growth due to increments in the number of body cells is more striking An increase in cell size and number accounts for growth of muscles during adolescence (Cheek and Hill, 1970) As seen in Figure 7 127, the adolescent growth spurt in boys is accompanied by an increase in number of muscle nuclei as compared to girls On the other hand, muscle cell size increases more rapidly in girls and by 10 years of age, the girl has attained maximal size for muscle cells The boy does not catch up until 15 years of age (Fig 7 128) Ovariectomy produces a pattern of muscle growth characteristic of the male in rats while adrenalectomy and castration of the male retard cell growth to some extent

Cell size and number are controlled by growth hormone, insulin, and thyroxine Rats that are hypophysectomized have reduced numbers of muscle cells (Table 7 19) Similarly, reduced numbers of nuclei are found in muscle of boys with pituitary insufficiency An increase in cell number and decrease in cell size result from injections of growth hormone Children with hypothyroidism also have re-

duced numbers of muscle cells and larger cells These changes can be reversed by treatment with thyroxine On the other hand, treatment with insulin does not decrease cell size In liver, both insulin and growth hormone increase weight, RNA, and protein, but only growth hormone produces a larger increase in deoxyribonucleic acid (DNA) These findings suggest that insulin is the key to cytoplasmic growth while growth hormone is required for nuclear proliferation

TURNOVER OF CELL PROTEIN

Cytoplasmic growth is determined by the turnover of cell protein Turnover is controlled by factors affecting both protein synthesis and degradation (Fig 7 129) Since half-lives of protein in mammalian cells (4 to 6 days) are far shorter than half-lives of the cells, changes in these rates will influence cell size, protein concentration, and the nature of protein Definitive demonstrations of continual protein turnover were not made until radioactive tracer methods became available in the late 1930's (Schoenheimer and Rittenberg, 1940, Schoenheimer, 1942) Radioactive amino acids have been used to investigate the effects of hormones on protein turnover as exemplified by studies in hypophysectomized rats using ^{14}C-leucine (Goldberg 1969a) Muscle protein was labeled by a single injection of tracer 48 hr prior to induction of muscle growth Retention of radioactivity in the protein is regulated by the rate of protein degradation while the fall in specific activity of prelabeled protein is controlled by the rate of protein synthesis Contrasting effects of work-induced and hormone-induced growth are shown in Figure 7 130 Both factors increase size of the soleus but only work-induced growth is associated with increased retention of radioactivity in muscle protein Both factors decrease the specific activity of muscle proteins These studies indicate that work-induced growth of the soleus involves both inhibition of protein degradation and stimulation of protein synthesis Growth hormone appears to affect only the rate of synthesis

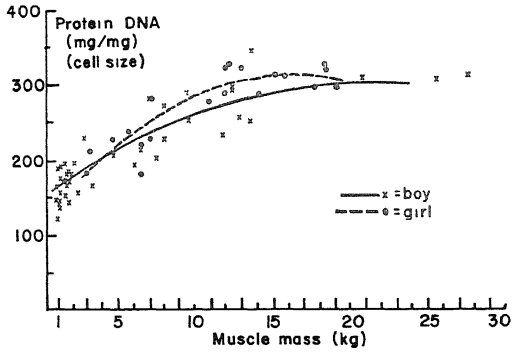


Fig 7 128 Muscle cell size (protein-to-DNA) is related to body size (muscle mass) for boys and girls during growth Note that growth by increments in cell size is less remarkable than by nuclear number increase Note also that females accelerate the growth of muscle cell size faster than boys although the latter eventually catch up (From Cheek and Hill 1970)

Decrease in muscle size occurs rapidly following disuse or denervation or following administration of large doses of adrenal steroids For atrophy to occur, protein degradation must exceed protein synthesis Decreased size of soleus and plantaris is seen following denervation or cortisone treatment (Fig 7 131) In each case, these factors facilitate degradation

Table 7 19
Hypophysectomized rats—effect of hormone therapy[a]

Animal	Muscle		Liver		
	Protein to DNA	No nuclei in muscle mass	Total protein	Total DNA	Protein to DNA
	mg /g	$N \times 10^9$	*mg*	*mg*	*mg /mg*
Intact	135	6 01			
Hypox	184	4 46	641	11 4	56 3
Hypox + GH	153	8 06	937	18 7	51 7
Hypox + Ins	175	5 29	903	13 2	68 7

From Cheek and Hill, 1970
[a] Cell size is indicated by the protein DNA ratio, and cell number by the nuclear count

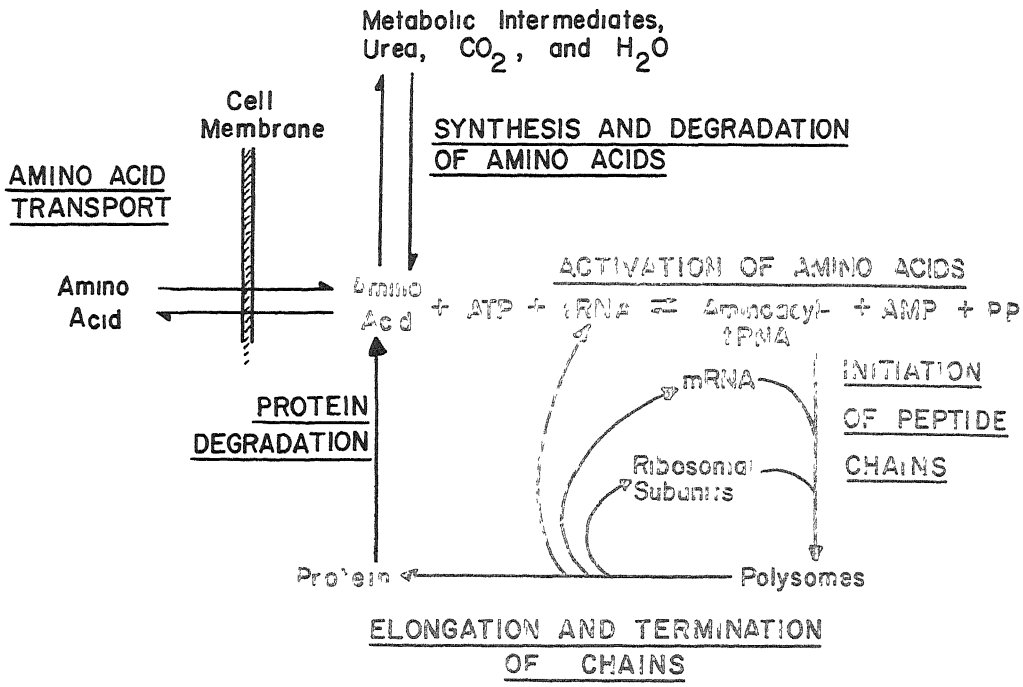


Fig 7 129 Steps involved in control of cell protein levels In this scheme synthesis of protein from the intracellular amino acid pool is considered to consist of the following steps *1)* activation of amino acids *2)* initiation of peptide chains on ribosomes and *3)* elongation and termination of peptide chains Amino acids may be added to this pool either by *1)* membrane transport *2)* intracellular synthesis from metabolites or *3)* protein degradation

of prelabeled protein Since the specific activity of the protein from denervated or cortisone-treated muscle is as high as the controls, inhibition of protein synthesis is indicated Inhibition of protein synthesis in steroid-treated animals may involve reduced rates of amino acid transport (Kostyo and Redmond, 1966) and of subsequent incorporation into protein (Wool and Weinshelbaum, 1959, Korner, 1967).

Insulin inhibits protein degradation in skeletal muscle and liver Decreased release of free amino acids from the forearm of postabsorptive man follows infusion of insulin into the brachial artery (Pozefsky et al , 1969) Insulin and free amino acids retard protein degradation in the perfused rat liver (Mortimore and Mondon, 1970, Woodside and Mortimore, 1970) Effects of the hormone on protein synthesis will be discussed later in this chapter

PROTEIN DEGRADATION

Protein degradation involves essentially all cell proteins, structural and enzymatic, and occurs within the cells Rates of degradation can vary markedly from one enzyme to another and can be altered by availability of substrate and other factors (Schimke, 1970) The rate is determined by properties of the protein as a substrate for a nonspecific degradative system as indicated by the first order relationship between rate and protein concentration Changes in the distribution of substrates and cofactors such as occur under various hormonal and nutritional conditions may lead to a variety of effects on specific enzymes to alter the rate of degradation Schimke et al (1964) provided the first clear example that substrate could modify the rate of degradation when tryptophan was found to retard degradation of tryptophan pyrrolase

The biochemical systems involved in protein degradation are poorly understood at present Lysosomes have been shown to digest phagocytized material as well as intracellular protein in viable cells (deDuve and Wattiaux, 1966, deReuck and Cameron, 1963) Increased acid phosphatase, a lysosomal enzyme, is present within the fibers of denervated muscle On the other hand, increased proteolysis in muscle of

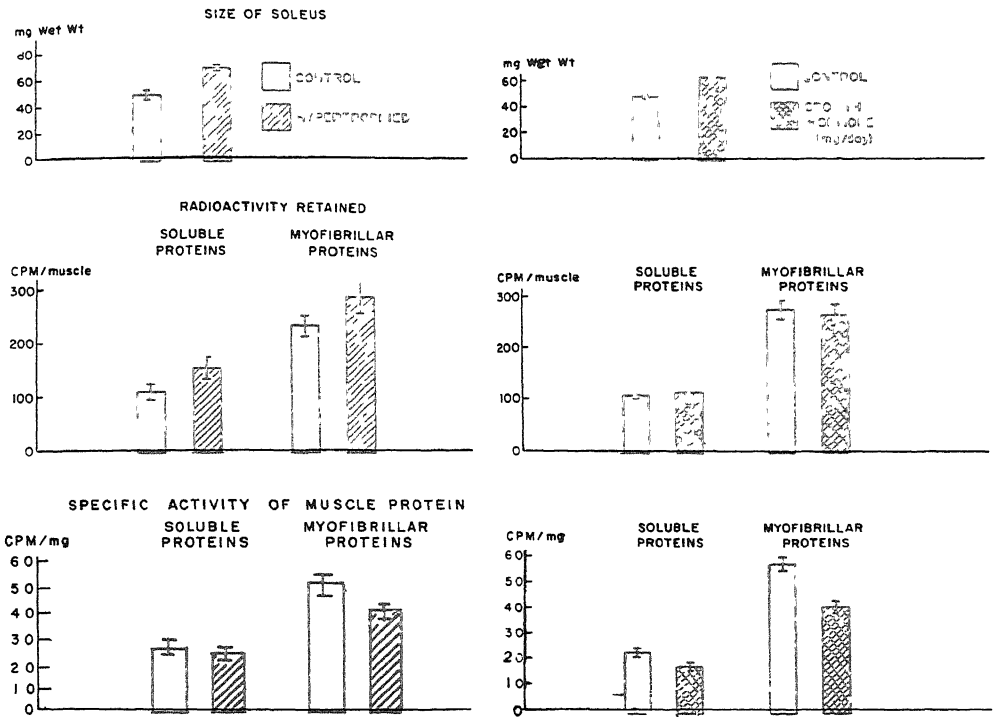


Fig 7 130 Effects of work hypertrophy and growth hormone on labeled proteins in soleus muscles of hypophysectomized rats Animals were sacrificed 6 days after unilateral tentomy of the gastrocnemius Results are compared with contralateral control Growth hormone was injected for 7 days and results compared to untreated hypophysectomized animals (From Goldberg 1969a)

cortisone-treated animals is not associated with increased activity of acid hydrolases (Buchanan and Schwartz, 1967) In contrast, these enzymes are raised in muscle extracts immediately following adrenalectomy These observations suggest that increased protein catabolism induced by denervation and cortisone is mediated by different mechanisms

PROTEIN SYNTHESIS

Protein synthesis may be regulated by factors affecting levels of structural components of the pathway, by supply of substrates, or by regulation of enzymes involved in protein synthesis Increased quantities of transfer RNA or ribosomes provide greater synthetic capacity while increased levels of messenger RNA may result in both qualitative and quantitative changes in the type of protein that is made Numbers of ribosomes are reduced in muscle and liver of hypophysectomized animals (Korner, 1964) Treatment of the animal with growth hormone restores these levels to normal Similarly, RNA content of muscle from diabetic animals is below normal and may contribute to the reduced rate of protein synthesis

In microorganisms, half-life of messenger RNA is in the range of seconds or minutes, but in animal cells, the template for protein synthesis persists for hours or days Therefore, increased levels of a specific protein can result from increased rates of translation of pre-existing mRNA or from stimulation of the synthesis of new mRNA. Effects of hydrocortisone and insulin on the levels of tyrosine α-ketoglutarate transaminase in cultured hepatoma cells provide examples of both mechanisms of induction (Lee et al , 1970) As seen in Figure 7.132, tyrosine transaminase activity increased approximately 10-fold following addition of hydrocortisone and approximately fourfold following addition of insulin. A lag of 1 hr is encountered before the steroid effect is evident, but the insulin effect begins immediately Addition of actinomycin D, an inhibitor

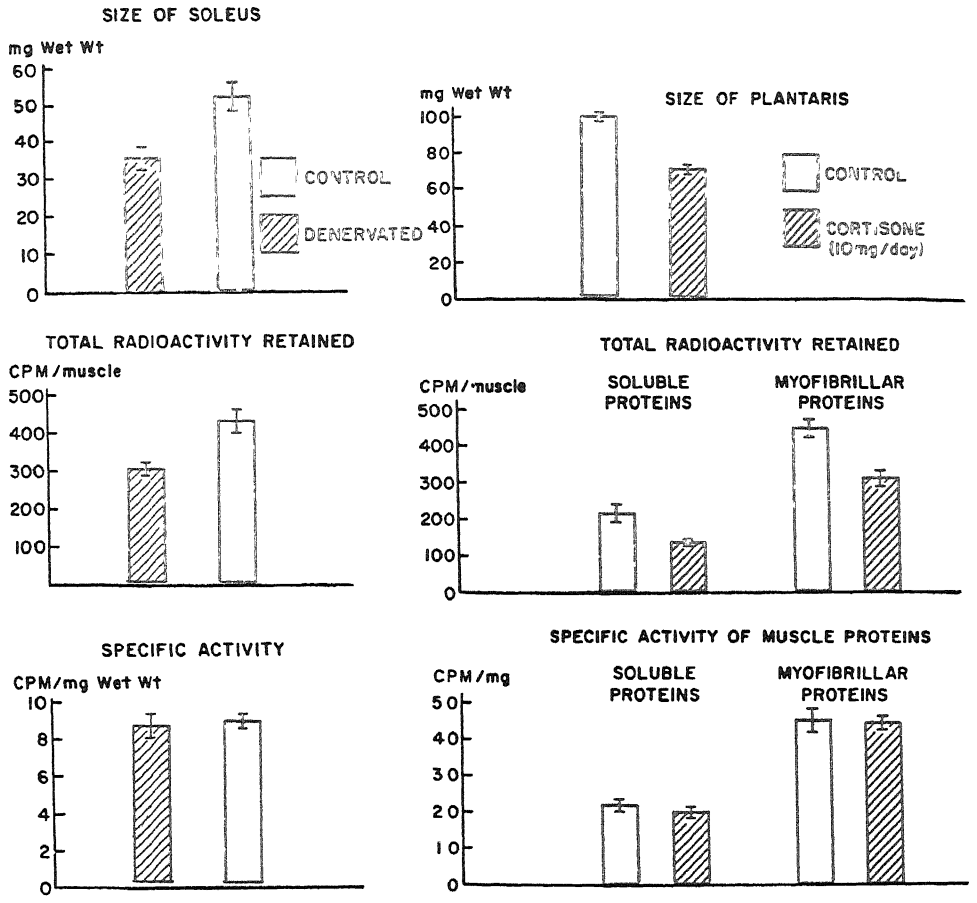


Fig 7 131 Effects of denervation and cortisone on labeled proteins in soleus and plantaris of hypophysectomized rats Animals were sacrificed 10 days after denervation Cortisone was administered for 7 days (From Goldberg 1969b)

of mRNA synthesis, along with the hormone completely blocks the effect of hydrocortisone but interfered with the insulin effect only after 2 hr of incubation These observations can be explained if hydrocortisone acts to increase enzyme levels by accelerating synthesis of mRNA within the nucleus while insulin facilitates reading of pre-existing mRNA An inhibitory effect of actinomycin in insulin-treated cells becomes apparent only when pre-existing mRNA has been degraded Half-life of tyrosine transaminase mRNA is estimated to be 2 to 3 hr In other experiments, effects of hydrocortisone and insulin were shown to be additive, strengthening the conclusion that these hormones act by different mechanisms

Cyclic-adenosine monophosphate (AMP) also functions as an inducer of tyrosine transaminase in liver cells (Wicks, 1969) This compound is proposed as the mediator of the effects of glucagon, epinephrine, and isoproterenol on tyrosine transaminase activity A comparison of the effectiveness of these inducers is presented in Table 7 20 Stimulation of enzyme synthesis by glucagon, isoproterenol, and dibutryl cyclic-AMP are inhibited by actinomycin D, suggesting that synthesis of new mRNA is involved in the effect Phosphorylation of nuclear histone by a histone kinase that is dependent upon cyclic-AMP may be involved in this induction (Langan, 1970) Activity of this kinase is increased fivefold by addition of cyclic-AMP resulting in phosphorylation of a specific serine residue in lysine-rich histone When glucagon was injected into rats, incorporation of radioactive phosphorus into liver histones was increased about 20-fold As a result of phosphorylation,

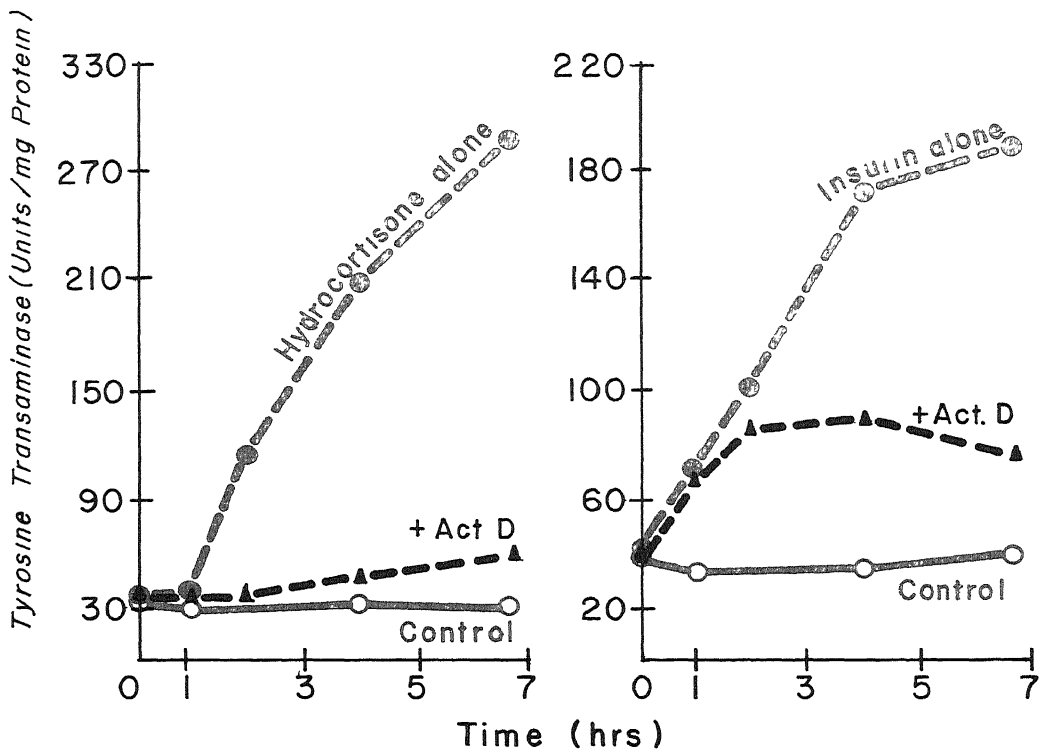


Fig 7 132 Induction of tyrosine αketoglutarate transaminase in cultured hepatoma cells by hydrocortisone and insulin Hydrocortisone concentration was 10^{-6}M and insulin concentration was 1 milliunit/ml Actinomycin D (0 2μg/ml) was added 30 min prior to addition of the hormone (From Lee et al 1970)

Table 7 20
Induction of tyrosine transaminase by hormones[a]

Additions	Final Concentration	Tyrosine Transaminase
	M	*units/mg protein*
None		7 5 ± 0 6
Hydrocortisone	1×10^{-6}	11 2 ± 0 7
Dibutyryl cyclic AMP	9×10^{-4}	18 4 ± 0 7
Glucagon	$1 1 \times 10^{-4}$	17 1 ± 1 0
Isoproterenol	$1 2 \times 10^{-4}$	15 6 ± 0 5
Insulin	$7 5 \times 10^{-5}$	9 5 ± 1 3

From Wicks, 1969

[a] An organ culture of fetal liver was used Additions were made after 42 hr in culture Two hours later the tissue was homogenized and tyrosine transaminase was assayed

histone may dissociate from specific regions of DNA, allowing for its transcription Effects of cyclic-AMP on mRNA synthesis in bacteria have also been observed (Varmus et al , 1970)

EFFECTS OF SUBSTRATE SUPPLY ON PROTEIN SYNTHESIS

A number of hormones including insulin, glucocorticoids, thyroxine, growth hormone, and estrogen increase entry of amino acids into various cells of the body Importance of these effects for the overall stimulation of protein synthesis depends upon the degree to which intracellular amino acid concentrations restrict protein synthesis in these cells In general, tissues, such as liver, that actively degrade amino acids to form glucose, fatty acids, and urea, are more dependent on a continued supply of amino acids from the plasma than are heart or skeletal muscle in which amino acid degradation is slow As seen in Figure 7 133, protein synthesis in liver varies from 60 to 200% of normal when amino acid concentrations in the perfusate range from 0 to 8 times normal plasma levels Tryptophan has been identified as an amino acid that is particularly important in maintaining normal rates of pro-

tein synthesis in this tissue (Munro, 1970) Low plasma levels of tryptophan result in low levels of tryptophan-tRNA, the activated form of the amino acid (Allen et al , 1969) On the other hand, synthesis of heart protein varies between 85 and 120% of the normal rate over this concentration range of plasma amino acids These findings indicate that a hormone-induced increase in amino acid entry into liver is of importance for the overall effect on protein synthesis, while this mechanism of action is of lesser importance for the overall effect in heart muscle

The effect of amino acid levels on protein synthesis is reflected in the aggregation of heart and liver ribosomes As shown in Figure 7 134, aggregation of liver ribosomes is restored to in vivo levels when livers are perfused with 5 times normal plasma levels of amino acids On the other hand, aggregation of heart ribosomes is only modestly responsive to variations in plasma amino acids Since these experiments are carried out in perfusion systems that lacked hormones, aggregation of liver ribosomes is more dependent upon the supply of amino acids than heart ribosomes Since incorporation of amino acids into protein increases as perfusate amino acid levels are raised, higher intracellular levels of amino acids accelerate initiation of peptide chains and results in increased numbers of polysomes and reduced levels of ribosomal subunits These findings indicate that increased amino acid supply can increase the rate of protein synthesis by raising intracellular levels of aminoacyl-tRNA and by accelerating initiation of peptide chains

REGULATION OF THE RIBOSOME CYCLE

Initiation, elongation, and termination of peptide chains make up the major groups of reactions involved in the cycling of ribosomal

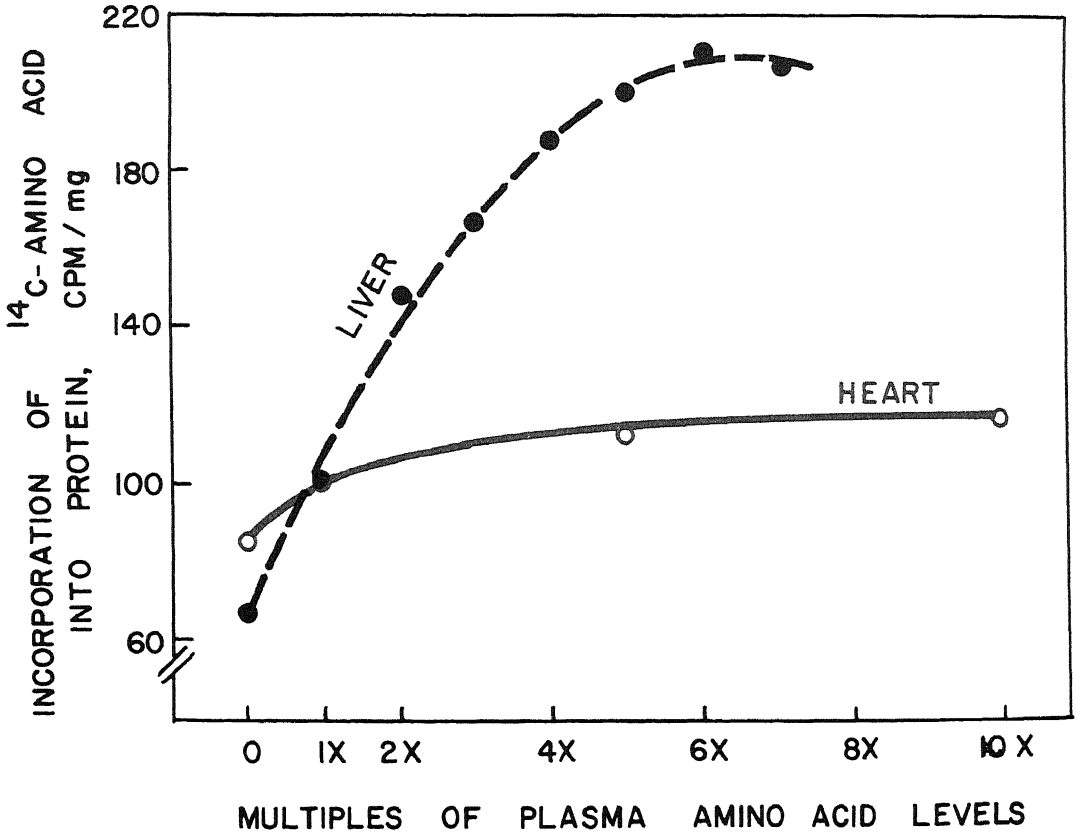


Fig 7 133 Effect or perfusate amino acid levels on incorporation of ^{14}C-amino acids into heart and liver protein Hearts and livers were perfused with buffer-containing multiples of normal plasma levels of amino acids Incorporation of ^{14}C-phenylalanine into heart protein and incorporation of ^{14}C-valine into liver protein were measured The zero point indicates that only the radioactive amino acid at normal plasma levels was added (From Jefferson and Morgan unpublished)

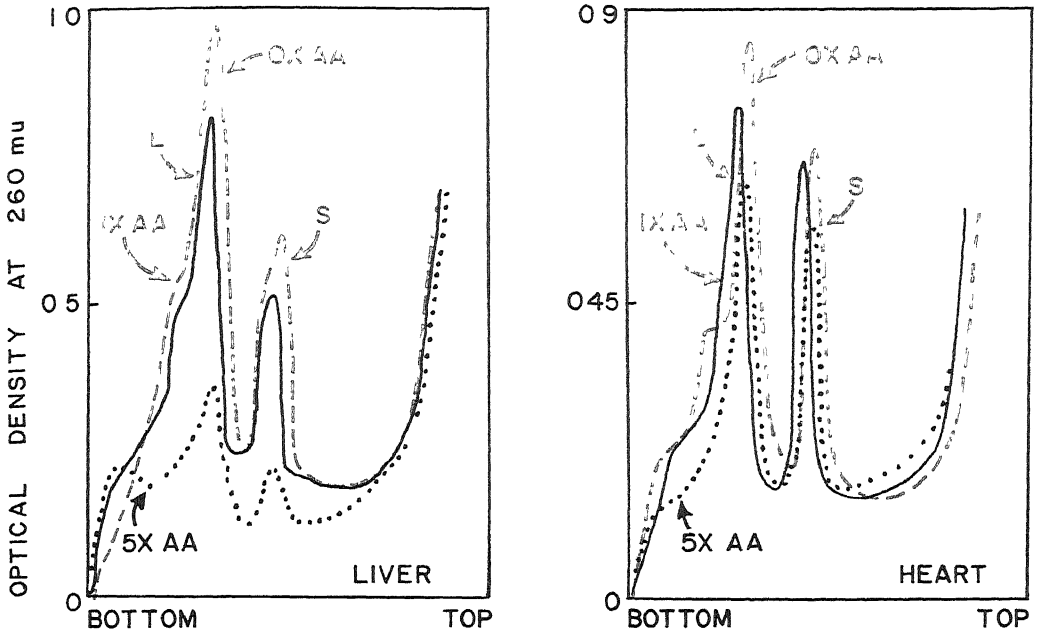


Fig 7 134 Effect of perfusate amino acid levels on aggregation of liver and heart ribosomes Hearts and livers were perfused with buffer-containing multiples of normal plasma levels of amino acids Aliquots of supernatants of homogenized liver and heart were layered on sucrose gradients and centrifuged Gradients were analyzed by pumping the contents of the tube through the flow cell of a recording spectrophotometer L indicates the large ribosomal subunit S the small subunit and 0 1 and 5 XAA the perfusate amino acid concentration in multiples of the normal plasma levels Aggregation of ribosomes in unperfused liver is similar to patterns of livers perfused with 5 times normal plasma levels of amino acids Aggregation of ribosomes in unperfused hearts is shown in Figure 7 135 (From Jefferson unpublished and Morgan et al 1971)

subunits into polysomes, down the length of messenger RNA, and back into the subunit pool If initiation of peptide chains is slow relative to elongation, ribosomal subunits accumulate and the rate of protein synthesis is slow On the other hand, inhibition of the elongation or termination groups of reactions results in decreased numbers of ribosomal subunits in association with a reduced rate of protein synthesis (Morgan et al , 1971) As indicated in Figure 7 134, a block in initiation of peptide chains develops in both heart and liver when these organs are perfused, in vitro, with normal plasma levels of amino acids In heart, this block persists even when high levels of amino acids are added.

Initiation of peptide chains in heart muscle is accelerated by addition of either insulin or fatty acids As shown in Figure 7 135, hearts that are perfused for 1 hr with buffer containing normal plasma levels of amino acids and glucose have fewer polysomes and more ribosomal subunits than unperfused hearts Addition of either insulin or fatty acid increases the number of polysomes and decreases subunits Faster rates of protein synthesis are also found These changes indicate that both of these factors accelerate initiation of peptide chains Since these effects are also demonstrable when 5 times normal plasma levels of amino acids are present, increased intracellular amino acid levels cannot account for the effect The mechanisms of the hormone and fatty acid effects on initiation are unknown, but may involve changes in intracellular levels of regulators of reactions involved in initiation The role of other hormones such as growth hormone, glucagon, and epinephrine in controlling steps in the ribosome cycle is not decided, but regulation at this level appears to be likely

Protein synthesis in skeletal muscle is controlled by insulin, growth hormone, glucocorticoids, and testosterone as well as the nutritional state of the animal As shown in Figure 7 136, aggregation of psoas muscle ribosomes is decreased when rats are fasted for 48 hr Perfusion of skeletal muscle with buffers containing

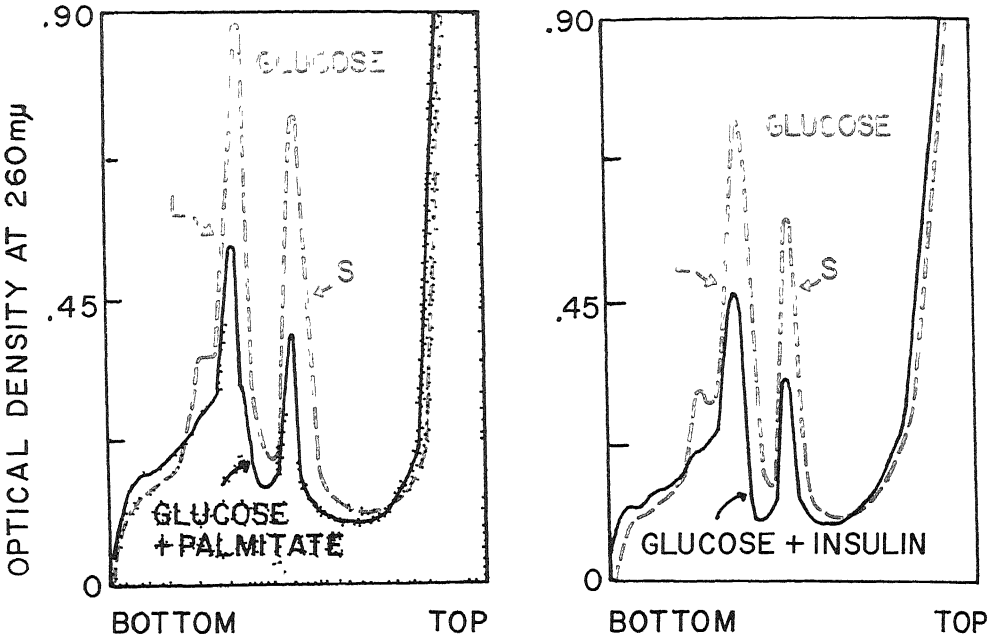


Fig 7 135 Effects of palmitate and insulin on aggregation of heart ribosomes Hearts were removed from rats fasted overight and perfused with buffer-containing glucose 3% serum albumin and normal plasma levels of all amino acids Insulin (25 mμ/ml) or palmitate (1 5m)) was added as indicated Aggregation of ribosomes was analyzed on sucrose density gradients as described in Figure 7 134 L indicates the large ribosome subunit S the small subunit and the shaded area the distribution of ribosomes in unperfused hearts Incorporation of ^{14}C-phenylalanine into heart protein was 94 c p m /mg when only glucose was present and 120 c p m /mg when either fatty acid or insulin was added (From Rannels et al 1970)

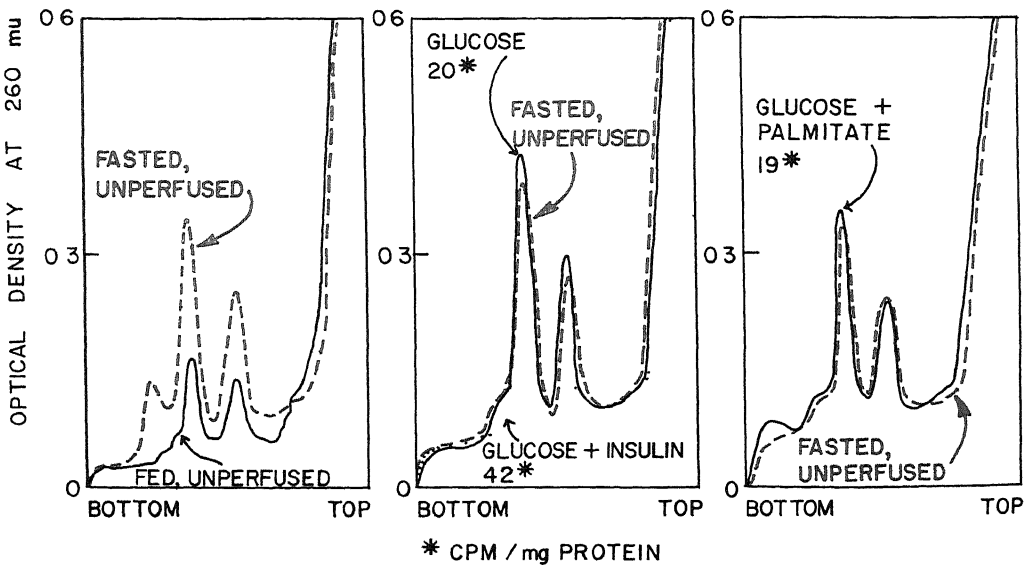


Fig 7 136 Effects of fasting, insulin and fatty acid on aggregation of psoas muscle ribosomes and on phenylalanine incorporation Psoas muscles were removed from fed or 48-hr fasted rats or from an isolated hemicorpus preparation following 90 min of perfusion with buffer-containing glucose glucose plus insulin or glucose plus palmitate Aggregation of ribosomes was analyzed on sucrose density gradients as described in Figure 7 134 The ^{14}C-phenylalanine incorporated into psoas muscle protein is shown for each of the muscles that were perfused (From Jefferson and Morgan unpublished)

insulin restores aggregation of ribosomes in muscle of starved animals, but fatty acids have no effect These changes are associated with increased rates of protein synthesis following insulin treatment, but no change in synthesis with fatty acid In comparison to heart muscle, control of the ribosome cycle in skeletal muscle by insulin is operative, but control by fatty acid is absent This situation provides for preservation of normal heart size during starvation and diabetes when insulin levels are low and fatty acid levels are high Utilization of skeletal muscle provides amino acids for glucose production through gluconeogenesis which is needed to support the metabolic needs of the brain

As indicated throughout this chapter, knowledge of hormonal control of growth and protein and nucleic acid metabolism is fragmentary Control of liver regeneration, cardiac hypertrophy, and atrophy and hypertrophy of skeletal muscle remains to be explained in terms of the control of protein synthesis and degradation Similarly, effects of hormones on the turnover of cell protein is only partially understood

BIBLIOGRAPHY

ABE K W E NICHOLSON G W LIDDLE, D N ORTH AND D P ISLAND Normal and abnormal regulation of β SMH in man *J Clin Invest* 48 1580–1585 1969a

ABE K, G A ROBISON G W LIDDLE R W BUTCHER W E NICHOLSON, AND C E BAIRD Role of cyclic AMP in mediating the effects of MSH, norepinephrine, and melatonin on frog skin color *Endocrinology* 85 674–682 1969b

ABEL J J Physiological, chemical and clinical studies on pituitary principles *Johns Hopkins Med J* 35 305–328, 1924

ABEL J J E M GEILING C A ROUILLER F K BELL AND O J WINTERSTEINER Crystalline insulin *J Pharmacol Exp Ther* 31 65–85, 1927

ADDISON T On disease of the suprarenal capsule *London Med Gaz* 43 517, 1855

ADIGA, P R R O HUSSA, M C ROBERTSON H R HOHL, AND T WINNICK Polysomes of bovine anterior pituitary gland and their role in hormone and protein biosynthesis *Proc Nat Acad Sci U S A* 60 603–613, 1968

AHLQUIST R P A study of the adrenotropic receptors *Amer J Physiol* 153 586–600 1948

AHREN, K, AND A HJALMARSON Early and late effects of growth hormone on transport of amino acids and monosaccharides in the isolated rat diaphragm In *Growth Hormone*, edited by A Pecile and E E Muller New York Excerpta Medica Foundation 1968 p 143–152

ALLEN R E, P L RAINES, AND D M REGEN Regulatory significance of transfer RNA charging levels I Measurements of charging levels in livers of chow fed rats, fasting rats, and rats fed balanced or imbalanced mixtures of amino acids *Biochim Biophys Acta* 190 323–336 1969

ANDERSON E, AND J A LONG The hormonal influences on the secretion of insulin *Recent Progr Hormone Res* 2 209–227, 1948

ANTON, A H AND D F SAYRE A study of the factors affecting the aluminum oxide-trihydroxyindole procedure for the analysis of catecholamines *J Pharmacol Exp Ther* 138 360–375 1962

ASHMORE, J, A B HASTINGS, F B NESBETT AND A E RENOLD Studies on carbohydrate metabolism in rat liver slices VI Hormonal factors influencing glucose 6 phosphatase *J Biol Chem* 218 77–88, 1956

ASHMORE J, AND D MORGAN Metabolic effects of adrenal glucocorticoid hormones In *The Adrenal Cortex*, edited by A B Eisenstein Boston Little, Brown 1967, p 249–267

AUGUST J T, D H NELSON AND G W THORN Response of normal subjects to larger amounts of aldosterone *J Clin Invest* 37 1549–1555, 1958

AURBACH, G D Isolation of parathyroid hormone after extraction with phenol *J Biol Chem* 234 3179–3181, 1959

AXELROD, J O-methylation of epinephrine and other catechols *in vitro* and *in vivo* *Science* 126 400–401 1957

BAIRD, D T, AND A GUEVARA Concentration of unconjugated estrone and estradiol in peripheral plasma in nonpregnant women throughout the menstrual cycle, castrate and postmenopausal women and in men *J Clin Endocr* 29 149–156, 1969

BAIRD, D T, R HORTON, C LONGCOPE, AND J F TAIT Steroid dynamics under steady state conditions *Recent Progr Hormone Res* 25 611–664 1969

BAISSET A AND P MONTASTRUC Polyurie par distension auriculaire chez le chien role de l hormone antidiuretique *J Physiol (Paris)* 49 33–36 1957

BAKER, T G A quantitative and cytological study of germ cells in human ovaries *Proc Roy Soc (London) [Biol] Ser B* 158 417–433 1963

BANTING F G AND C H BEST Internal secretion of pancreas *J Lab Clin Med* 7 251–266 1922

BARDIN C W AND J A MAHOUDEAU Dynamics of androgen metabolism in women with hirsutism *Ann Clin Res* 2 251–262 1970

BARGMANN W Uber die neurosekretorische Verknupfung von Hypothalamus und Neurohypophyse *Z Zellforsch* 34 610–634, 1949

BARGMANN W *Das Zwischenhirn—Hypophysensystem* Berlin Springer 1954

BARKER, S B Physiological activity of thyroid hormones and analogues In *The Thyroid Gland* edited by R Pitt Rivers and W R Trotter London Butterworths, 1964, vol 1, p 199–236

BARR M L, AND E G BERTRAM A morphological distinction between neurones of the male and female and the behavior of the nucleolar satellite during accelerated nucleoprotein synthesis *Nature (London)* 163 676–677 1949

BARR, M L, D H CARR, J POZSONYI R A WILSON H G DUNN, T S JACOBSON, J R MILLER, M LEWIS AND B CHOWN The XXXXY sex chromosome abnormality *Canad Med Ass J* 87 891–901 1963

BARRACLOUGH C A Modifications in the CNS regulation of reproduction after exposure of prepubertal rats to steroid hormones *Recent Progr Hormone Res* 22 503–539 1966

BASHIRELAHI N, G J CHADER, AND C A VILLEE Effects of dihydrotestosterone on the synthesis of nucleic acid and ATP in prostate nuclei *Biochem Biophys Res Commun* 37 976–981 1969

BATES, R W, AND P G CONDLIFFE The physiology and chemistry of thyroid stimulating hormone In *The Pituitary Gland Anterior Pituitary* edited by G W Harris and B T Donovan Berkeley Univ California Press 1966, vol 1 p 374–410

BAULIEU, E E, F PEILLON, AND C J MIGEON Adrenogenital syndrome In *The Adrenal Cortex*, edited by A B Eisenstein Boston Little, Brown 1967, p 553–637

BAYARD F I Z BEITINS, A KOWARSKI, AND C J MIGEON Measurement of plasma aldosterone by radioimmunoassay *J Clin Endocr* 31 1–6 1970

BAYLIS, E M, F GREENWOOD, V JAMES, J JENKINS J LANDON, V MARKS, AND E SAMOLS An examination of the control mechanisms postulated to control growth hormone secretion in man In *Growth Hormone*, edited by A Pecile and E E Muller New York Excerpta Medica Foundation 1968 p 89–104

BELL, N H Effects of glucagon, dibutryl cyclic 3′ 5′ adenosine monophosphate, and theophylline on calcitonin *in vitro* *J Clin Invest* 49 1368–1373, 1970

BENIRSCHKE K Adrenals in anencephaly and hydrocephaly *Obstet Gynecol* 8 412–425 1956

BERNICK S AND B H ERSHOFF Histochemical study of bone in cortisone treated rats *Endocrinology* 72 231–237 1963

BERSON S A AND R S YALOW Immunoassay of protein hormones In *The Hormones* edited by G Pincus K V Thimann and E B Astwood New York Academic Press 1964 vol 4 p 557–630

BERSON S A AND R S YALOW Radioimmunoassays of peptide hormones in plasma *New Eng J Med* 277 640–647 1967

BERSON S A AND R S YALOW Radioimmunoassay of ACTH in plasma *J Clin Invest* 47 2725–2751 1968

BESSER G M D N ORTH W E NICHOLSON, R L BYYNY K ABE AND J P WOODHAM Dissociation of the disappearance of bioactive and radioimmunoreactive ACTH from plasma in man *J Clin Endocr* 1971 In press

BEYER C AND C H SAWYER Hypothalamic unit activity related to control of the pituitary gland *Frontiers in Neuroendocrinology* edited by W F Ganong and L Martini New York Oxford Univ Press 1969 p 255–287

BEYER K F AND L T SAMUELS Distribution of steroid 3β ol dehydrogenase in cellular structures of the adrenal gland *J Biol Chem* 219 69–76 1956

BIGLIERI, E G 17 Hydroxylation deficiency in man *J Clin Invest* 45 987, 1966

BISHOP, P M Dr Addison and his work *Proc Roy Soc Med* 48 1032–1038 1955

BLANDAU R Interval secretion In *Sex and Internal Secretions* edited by W C Young Baltimore Williams & Wilkins 1962, vol 1 p 797–882

BLIZZARD R M, AND M KYLE Studies of the adrenal antigens and antibodies in Addison s disease *J Clin Invest* 42 1653–1660 1963

BLIZZARD R M G W LIDDLE, C J MIGEON AND L WILKINS Aldosterone excretion in virilizing adrenal hyperplasia *J Clin Invest* 38 1442–1451 1959

BLOCH, E, AND K BENIRSCHKE Synthesis *in vitro* of steroids by human fetal adrenal gland slices *J Biol Chem* 234 1085–1089 1959

BONDY P K, F L ENGEL, AND B W FARRAR The metabolism of amino acids and protein in the adrenalectomized nephrectomized rat *Endocrinology* 44 476–483, 1949

BONGIOVANNI A M The adrenogenital syndrome with deficiency of 3β hydroxy steroid dehydrogenase *J Clin Invest* 41 2086–2092 1962

BONGIOVANNI, A M AND G W CLAYTON JR A simplified method for the routine determination of pregnanediol and pregnane triol in urine *Johns Hopkins Med J* 94 180–186 1954

BORKOWSKI, A J S LEVIN, C DELCROIX, A MAHLER, AND V VERHAS Blood cholesterol and hydrocortisone production in man quantitative aspects of the utilization of circulating cholesterol by the adrenals at rest and under adrenocorticotropin stimulation *J Clin Invest* 46 797–811, 1967

BOUCHER, R, R VEYRAT J DE CHAMPLAIN, AND J GENEST New procedures for measurement of human plasma angiotensin and renin activity levels *Canad Med Ass J* 90 194–201, 1964

BOWERS, C Y, A V SCHALLY F ENZMANN, J BOLER, AND K FOLKERS Porcine thyrotropin releasing hormone is (pyro)glu his pro(NH_2) *Endocrinology* 86 1143–1153 1970a

BOWERS C Y, A V SCHALLY D S SCHALCH, C GUAL, A KASTIN E CASTANEDA, AND K FOLKERS Synthetic thyrotropin releasing hormone (TRH) effect in man (Program of Endocrine Society) *Endocrinology* 86 1188, 1970b

BOYD G W, J LANDON, AND W S PEART Radioimmunoassay for determining plasma levels of angiotensin II in man *Lancet* 2 1002–1005, 1967

BOGGS D R, J W ATHENS, G E CARTWRIGHT, AND M M WINTROBE Leukokinetic studies IX Experimental evaluation of a model of granulopoiesis *J Clin Invest* 44 643–656, 1965

BRANSOME, E D, JR Adrenal cortex *Ann Rev Physiol* 30 171–212, 1968

BRESSLER, R The biochemistry of ketosis *ANN N Y Acad Sci* 104 735–752, 1963

BREWER, H B, JR, AND R RONAN Bovine parathyroid hormone amino acid sequence *Proc Nat Acad Sci USA* 67 1862–1869, 1970

BROMER W W L G SINN A STAUB AND O K BEHRENS The amino acid sequence of glucagon *Diabetes* 6 234–238 1957

BROWN R D C A STROTT AND G W LIDDLE Site of stimulation of aldosterone biosynthesis by angiotensin and potassium *Clin Res* 19 32 1971

BROWN GRANT K The control of TSH secretion In *The Pituitary Gland Anterior Pituitary* edited by G W Harris and B T Donovan Berkeley Univ California Press 1966 vol 2 p 235–269

BRUBACHER F S AND A J VANDER Sodium deprivation and renin secretion in unanesthetized dogs *Amer J Physiol* 214 15–21 1968

BRUCHOVSKY, N, AND J D WILSON The intranuclear binding of testosterone and 5-α androstan 17β ol 3 one by rat prostate *J Biol Chem* 243 5953–5960 1968

BUCHANAN W E AND T B SCHWARTZ Lysosomal enzyme activity in heart and skeletal muscle of cortisone treated rats *Amer J Physiol* 212 732–736 1967

BURGUS R T F DUNN D DESIDERIO D N WARD W VALE AND R GUILLEMIN Characterization of ovine hypothalamic hypophysiotropic TSH releasing factor *Nature (London)* 226 321–325, 1970

BURGUS R, AND R GUILLEMIN Chemistry of thyrotropin releasing factor (TRF) In *Hypophysiotropic Hormones of the Hypothalamus Assay and Chemistry* edited by J Meites Baltimore Williams & Wilkins 1970a p 227–241

BURGUS, R AND R GUILLEMIN Hypothalamic releasing factors *Ann Rev Biochem* 39 499–526 1970b

BUTCHER, R W C E BAIRD AND E W SUTHERLAND Effects of lipolytic and antilipolytic substances on adenosine 3 5′ monophosphate levels in isolated fat cells *J Biol Chem* 243 1705–1712 1968

CAHILL G F, JR, E B MARLISS, AND T T AOKE Fat and nitrogen metabolism in fasting man In *Adipose Tissue Regulation and Metabolic Function* edited by B Jeanrenaud and D Hepp New York Academic Press 1970, p 181–185

CALDEYRO BARCIA, R *Factors Controlling the Actions of the pregnant Human Uterus* Transactions of the 5th Macy Conference on Physiology Prematurity 1961 p 11–117

CALLOW, N H, R K CALLOW AND C W EMMENS Colorimetric determination of substances containing the grouping $—CH_2CO—$ in urine extracts as an indication of androgen content *Biochem J* 32 1312–1331 1938

CAMACHO A M AND C J MIGEON Testosterone excretion and production rate in normal adults and in patients with congenital adrenal hyperplasia *J Clin Endocr* 26 893–896 1966

CAMERINI DAVALOS R A AND COLE, H S *Early Diabetes Supplement 1, Advances in Metabolic Disorders* New York Academic Press, 1970

CARE A D, L M SHERWOOD J T POTTS, JR, AND G D AURBACH Evaluation by immunoassay of factors controlling the secretion of parathyroid hormone Nature (London) 209 52–55 1966

CASPI E, R I DORFMAN, B T KHAN, G ROSENFELD, AND W SCHMID Degradation of corticosteroids VI Origin of the carbon atoms of steroid hormones biosynthesized *in vitro* in bovine adrenal from acetate 1 ^{14}C *J Biol Chem* 237 2085–2088, 1962

CASTLES, T R AND H E WILLIAMSON Mediation of aldosterone induced antinatriuresis via RNA synthesis *Proc Soc Exp Biol Med* 124 717–719, 1967

CATT K J M C CAIN, AND J P COGHLAN Measurement of angiotensin II in blood *Lancet* 2 1005–1007, 1967

CHANNING C P Studies on tissue culture of equine ovarian cell types pathways of steroidogenesis *J Endocr* 43 403–414, 1969

CHASE L R, AND G D AURBACH Cyclic AMP and the mechanism of action of parathyroid hormone In *Parathyroid Hormone and Calcitonin*, edited by R V Talmage and L F Belanger New York Excerpta Medica Foundation, 1968, p 247–257

CHASE, L R, AND G D AURBACH The effect of parathyroid hormone on the concentration of adenosine 3′ 5′ monophosphate in skeletal tissue *in vitro* *J Biol Chem* 245 1520–1526, 1970

CHEEK, D B, AND D E HILL Muscle and liver cell growth role of hormones and nutritional factors *Fed Proc* 29 1503–1509 1970

CHRISTENSEN, H N, A M CULLEN, L I HARRISON, S S FAJANS, AND S PEK Reciprocal transport interactions between amino acids and pancreatic hormones *Fed Proc* 30 200, 1971

CONN J W Primary aldosteronism a new clinical syndrome *J Lab Clin Med* 45 3-17 1955

COPP D H Parathyroid hormone calcitonin and calcium homeostasis In *Parathyroid Hormone and Thyrocalcitonin* edited by R V Talmage and L F Belanger New York Excerpta Medica Foundation 1968 p 25-39

COPP D H Endocrine control of calcium homeostasis *J Endocr* 43 137-161 1969

COPP D H Endocrine regulation of calcium metabolism *Ann Rev Physiol* 32 61-86 1970

COPP D H E C CAMERON B A CHENEY A G F DAVIDSON AND K G HENZE Evidence for calcitonin—a new hormone from the parathyroid that lowers blood calcium *Endocrinology* 70 638-649 1962

COPP D H AND A G F DAVIDSON Direct humoral control of parathyroid function in the dog *Proc Soc Exp Biol Med* 107 342-344 1961

COPPAGE W S JR D P ISLAND A E COONER AND G W LIDDLE The metabolism of aldosterone in normal subjects and in patients with hepatic cirrhosis *J Clin Invest* 41 1672-1680 1962

CORBIN J D AND E G KREBS A cyclic AMP stimulated protein kinase in adipose tissue *Biochem Biophys Res Commun* 36 328-336 1969

CORI C F Mammalian carbohydrate metabolism *Physiol Rev* 11 143-275 1931

CRANE R K AND A SOLS The association of hexokinase with particulate fractions of brain and other tissue homogenates *J Biol Chem* 203 273-292 1953

CRIGLER J F S H SILVERMAN AND L WILKINS Further studies on the treatment of congenital adrenal hyperplasia with cortisone IV Effect of cortisone and compound B in infants with disturbed electrolyte metabolism *Pediatrics* 10 397-413 1952

CROSS B A Neural control of oxytocin secretion In *Neuroendocrinology* edited by L Martini and W F Ganong New York Academic Press 1966 vol 1 p 217-259

CUATRECASAS P Interaction of insulin with the cell membrane—the primary action of insulin *Proc Nat Acad Sci USA* 63 450-457 1969

CUSHING H The basophil adenomas of the pituitary body and their clinical manifestations (pituitary basophilism) *Johns Hopkins Med J* 50 137-195 1932

DANFORTH W H, AND E HELMREICH Regulation of glycolysis in muscle *J Biol Chem* 239 3133-3138 1964

DAUGHADAY W H The binding of corticosteroids by plasma protein In *The Adrenal Cortex*, edited by A B Eisenstein Boston Little Brown, 1967 p 385-403

DAVID M A, F FRASCHINI, AND L MARTINI Control of LH secretion role of a short feedback mechanism *Endocrinology* 78 55-60 1966

DAVIDSON, J M Feedback control of gonadotropin secretion In *Frontiers in Neuroendocrinology* edited by W F Ganong and L Martini New York Oxford Univ Press, 1969, p 343-388

DEBODO, R C R STEELE, N ALTSZULER A DUNN, AND J A BISHOP On the hormonal regulation of carbohydrate metabolism studies with C^{14} glucose *Recent Progr Hormone Res* 19 445-488 1963

DECRECCHIO, L Sopra un caso di apparenze virili in una donna *Il Morgagni* 7 151 1865

DE DUVE C AND R WATTIAUX Functions of lysosomes *Ann Rev Physiol* 28 435-492 1966

DELORES ARNAIZ, G R, AND E DEROBERTIS Cholinergic and non cholinergic nerve endings in the rat brain II *J Neurochem* 9 503-518, 1962

DELUCA, H F The function of vitamine D and its relationship to parathyroid hormone and calcitonin In *Calcitonin 1969* New York Springer Verlag, 1970 p 205-219

DEMOOR, P O STEENO M RASKIN, AND A HENDRIKX Fluorimetric determination of free plasma 11-hydroxycorticosteroids in man *Acta Endocr (Kobenhavn)* 33 297-307, 1960

DENTON, R M, AND B R MARTIN Pathways of lipid synthesis and their regulation in rat epididymal fat cells In *Adipose Tissue, Regulation and Metabolic Function Hormone and Metabolic Research, Supplement 2*, edited by B Jeanrenaud and D Hepp New York Academic Press, 1970, p 143-151

DERELCK A V S AND M P CAMERON (eds) *Lysosomes* Boston Little Brown 1963

DEXTER R N I M FISHMAN R L NEY AND G W LIDDLE An effect of adrenocorticotrophic hormone on adrenal cholesterol accumulation *Endocrinology* 81 1185-1187 1967

DOBYNS B M The thyroid pituitary relationships In *The Thyroid* edited by J B Hazard and D E Smith Baltimore Williams & Wilkins 1964 p 76-99

DOLE V P Energy storage In *Adipose Tissue Handbook of Physiology* edited by A E Renold and G F Cahill Jr Washington D C Amer Physiol Soc 1965 sect 5 p 13-18

DORFMAN R I D C SHARMA A L SOUTHREN AND J L GABRILOVE Biosynthesis of steroids in various tissues related to feminizing syndromes *Cancer Res* 25 1125-1128 1965

DOUGHERTY T F Role of steroids in regulation of inflammation In *Inflammation and diseases of connective tissue* edited by L C Mills and J H Moyer Philadelphia Saunders 1961 p 449-460

DOUGLAS W W Calcium dependent links in stimulus secretion coupling in the adrenal medulla and neurohypophysis In *Mechanism of Release of Biogenic Amines* edited by U S von Euler S Rosell and B Uvnas Oxford Pergamon 1966a p 267-290

DOUGLAS W W The mechanism of release of catecholamines from the adrenal medulla *Pharmacol Rev* 18 471-480 1966b

DOUGLAS W W Stimulus secretion coupling the concept and clues from chromaffin and other cells *Brit J Pharmacol* 34 451-474 1968

DOUGLAS W W AND A M POISNER Calcium movement in the neurohypophysis of the rat and its relation to the release of vasopressin *J Physiol (London)* 172 19-30 1964

EBERLEIN W R AND A M BONGIOVANNI Steroid metabolism in salt losing form of congenital adrenal hyperplasia *J Clin Invest* 37 889-890 1958

EDELMAN I S R BOGOROCH AND G A PORTER Specific action of aldosterone on RNA synthesis *Trans Ass Amer Physicians* 77 307-316 1964

EISENBERG E Renal effects of parathyroid hormone In *Parathyroid Hormone and Thyrocalcitonin* edited by R V Talmage and L F Belanger New York Excerpta Medica Foundation 1968 p 465-474

ENGEL F L A consideration of the roles of the adrenal cortex and stress in the regulation of protein metabolism *Recent Progr Hormone Res* 6 277-308 1950

ENGEL, F L AND T T AMATRUDA, JR Hormonal aspects of ketosis *Ann N Y Acad Sci* 104- 753-771 1963

ENGLAND, P J, AND P J RANDLE Effectors of rat heart hexokinase and the control of rates of glucose phosphorylation in the perfused rat heart *Biochem J* 105 907-920 1967

EWALD W WERBIN AND I L CHAIKOFF Δ^5 3 ketosteroid isomerase solubilization and stabilization in beef adrenal gland preparation *Steroids* 3 505-522 1964

EXTON, J H L E MALLETTE, L S JEFFERSON, E H A WONG N FRIEDMANN, T B MILLER JR AND C R PARK The hormonal control of hepatic gluconeogenesis *Recent Progr Hormone Res* 26 411-457 1970

EXTON J H, AND C R PARK Control of gluconeogenesis in liver I General features of gluconeogenesis in the perfused livers of rats *J Biol Chem* 242 2622-2636, 1967

EXTON, J H, AND C R PARK Control of gluconeogenesis in liver II Effects of glucagon catecholamines and adenosine 3 5 mono phosphate on gluconeogenesis in the perfused rat liver *J Biol Chem* 243 4189-4196 1968

EXTON J H, AND C R PARK Control of gluconeogenesis in liver III Effects of L lactate pyruvate fructose glucagon, epinephrine, and adenosine 3 5 monophosphate on gluconeogenic intermediates in the perfused rat liver *J Biol Chem* 244 1424-1433, 1969

FANESTIL, D D AND I S EDELMAN Characteristics of the renal nuclear receptors for aldosterone *Proc Nat Acad Sci USA* 56 872-879, 1966

FANESTIL, D D G A PORTER, AND I S EDELMAN Aldosterone stimulation of sodium transport *Biochim Biophys Acta* 135 74-88, 1967

FANG, S, K M ANDERSON, AND S LIAO Receptor proteins for androgens—on the role of specific proteins in selective retention of 17β hydroxy 5α-androstan 3-one by rat ventral prostate *in vivo* and *in vitro J Biol Chem* 244 6584-6595 1969

FARQUHAR M G Origin and fate of secretory granules in cells of the anterior pituitary gland *Trans N Y Acad Sci* 23 346-351 1961

FAWCETT, C P The present status of the chemistry of PIF FRF and LRF In *Hypophysiotropic Hormones of the Hypothalamus Assay and Chemistry* edited by J Meites Baltimore The Williams & Wilkins

FAWCETT D W J A LONG AND A L JONES The ultrastructure of endocrine glands *Recent Progr Hormone Res* 25 315-368 1969

FELIG P T POZEFSKY E MARLISS AND G F CAHILL JR Alanine key role in gluconeogenesis *Science* 167 1003-1004 1970

FERIN M A TEMPONE P E ZIMMERING AND R L VANDE WIELE Effect of antibodies to 17 β estradiol and progesterone on the estrous cycle of the rat *Endocrinology* 85 1070-1078 1969

FIELD J B Studies on the mechanism of action of thyroid stimulating hormone *Metabolism* 17 226-245 1968

FINN A L, AND L G WELT Effect of aldosterone administration on electrolyte excretion and glomerular filtration rate in the rat *Amer J Physiol* 204 243-244 1963

FISHER, C W R INGRAM AND S W RANSON *Diabetes Insipidus and the Neuro Hormonal Control of Water Balance* Ann Arbor Mich Edwards 1938a p 212

FISHER, C H W MAGOUN AND S W RANSOM Dystocia in diabetes insipidus The relation of pituitary oxytocin to parturition *Amer J Obstet Gynec* 36 1-9 1938b

FITZPATRICK, R J Blood concentrations of oxytocin in labour *J Endocr* 22 19-20 1961a

FITZPATRICK R J The estimation of small amounts of oxytocin in blood In *Oxytocin* edited by R Caldeyro Barcia and H Heller Oxford Pergamon Press 1961b p 358-377

FLATT J P Energy metabolism and the control of lipogenesis in adipose tissue In *Adipose Tissue Regulation and Metabolic Function Hormone and Metabolic Research Supplement 2* edited by B Jeanrenaud and D Hepp New York Academic Press, 1970 p 93-100

FLEISCHER N, AND W VALE Inhibition of vasopressin induced ACTH release from the pituitary by glucocorticoids *in vitro Endocrinology* 83 1232-1236 1968

FOLKERS K F ENZMANN, J BOLER, C Y BOWERS, AND A V SCHALLY Discovery of modification of the synthetic tripeptide sequence of the thyrotropin releasing hormone having activity *Biochem Biophys Res Commun* 37 123-126 1969

FORD C E K W JONES O J MILLER U MITTWOCH, L S PENROSE, M RIDLER, AND A SHAPIRO The chromosomes in a patient showing both mongolism and the Klinefelter syndrome *Lancet* 1 709-710, 1959

FORD, J K AND R W YOUNG Cell proliferation and displacement in the adrenal cortex of young rats injected with tritiated thymidine *Anat Rec* 146 125-133 1961

FRASER, D R AND E KODICEK Unique biosynthesis of kidney of a biologically active vitamin D metabolite *Nature (London)* 228 764-766, 1970

FREINKEL N The intermediary metabolism of thyroid tissue In *The Thyroid Gland*, edited by R Pitt Rivers and W R Trotter London Butterworths, 1964, vol 1, p 131-162

FUXE, K AND T HOKFELT Catecholamines in the hypothalamus and the pituitary gland In *Frontiers in Neuroendocrinology*, edited by W F Ganong and L Martini New York Oxford Univ Press, 1969, p 47-96

GANONG, W F AND L MARTINI (eds) *Frontiers in Neuroendocrinology* New York Oxford Univ Press, 1969

GAUER, O H, AND J P HENRY Circulatory basis of fluid volume control *Physiol Rev* 43 423-481, 1963

GAY, V L G D NISWENDER, AND A R MIDGLEY Response of individual rats and sheep to one or more injections of hypothalamic extract as determined by radioimmunoassay of plasma LH *Endocrinology* 86 1305-1312 1970

GEMZELL C, AND P ROOS The physiology and chemistry of follicle stimulating hormone In *The Pituitary Gland*, edited by G W Harris and B T Donovan Berkeley Univ California Press, 1966, vol 1, p 492-517

GESCHWIND, I I Mechanism of action of releasing factors In *Frontiers in Neuroendocrinology*, edited by W F Ganong and L Martini New York Oxford Univ Press, 1969, p 389-431

GESCHWIND I I Mechanism of action of hypothalamic adenohypophysiotropic factors In *Hypophysiotropic Hormones of the Hypothalamus Assay and Chemistry* edited by J Meites Baltimore Williams & Wilkins 1970 p 298-319

GILL G N AND L D GARREN On the mechanism of action of adrenocorticotropic hormone the binding of cyclic 3 5 adenosine monophosphate to an adrenal cortical protein *Proc Nat Acad Sci US* 4 63 512-518 1969

GLICK S M The regulation of growth hormone secretion In *Frontiers in Neuroendocrinology* edited by W F Ganong and L Martini New York Oxford Univ Press 1969 p 141-182

GLINSMANN W H AND G E MORTIMORE Influence of glucagon and 3 5 AMP on insulin responsiveness of the perfused rat liver *Amer J Physiol* 215 553-559 1968

GOLDBERG A L Protein turnover in skeletal muscle I Protein catabolism during work induced hypertrophy and growth induced with growth hormone *J Biol Chem* 244 3217-3222 1969a

GOLDBERG A L Protein turnover in skeletal muscle II Effects of denervation and cortisone on protein catabolism in skeletal muscle *J Biol Chem* 244 3223-3229 1969b

GOODMAN H M AND E KNOBIL Some endocrine factors in regulation of fatty acid mobilization during fasting *Amer J Physiol* 201 1 1961

GOOLDEN A W G Radioactive isotopes of iodine and their applications In *The Thyroid Gland* edited by R Pitt Rivers and W R Trotter London Butterworths 1964 vol 1 p 390-436

GORDON R E O KUCHEL, G W LIDDLE AND D P ISLAND Role of the sympathetic nervous system in regulating renin and aldosterone production in man *J Clin Invest* 46 599-605, 1967

GORSKI J D TOFT, G SHYAMALA D SMITH, AND A NOTIDES Hormone receptors studies on the interaction of estrogen with the uterus *Recent Progr Hormone Res* 24 45-80 1968

GRAHAME SMITH, D G R W BUTCHER, R L NEY, AND E W SUTHERLAND Adenosine 3 5 monophosphate as the intracellular mediator of the action of adrenocorticotropic hormone on the adrenal cortex *J Biol Chem* 242 5535-5541 1967

GRANT, N Metabolic effects of adrenal glucocorticoid hormones In *The Adrenal Cortex*, edited by A B Eisenstein Little Brown and Company, Boston Little, Brown, 1967 p 269-292

GREER, M A J W KENDALL AND M SMITH Antithyroid compounds In *The Thyroid Gland*, edited by R Pitt Rivers and W R Trotter London Butterworths, 1964, vol 1 p 357-389

GRODSKY G H LANDAHL, D CURRY, AND L BENNETT A two compartment model for insulin secretion In *Early Diabetes Supplement I, Advances in Metabolic Disorders* edited by R A Camerini Davalos and H S Cole New York Academic Press 1970, p 45-50

GROSS, J AND R PITT-RIVERS The identification of 3,4,4' triiodothyronine in human plasma *Lancet* 1 439, 1952

GRUMBACH M M AND M L BARR Cytologic tests of chromosomal sex in relation to sexual anomalies in man *Recent Progr Hormone Res* 14 255-334, 1958

GRUMBACH M M, S L KAPLAN, J J SCIARRA, AND I M BURR Chorionic growth hormone prolactin (CGP) secretion disposition, biologic activity in man, and postulated function as the "growth hormone" of the second half of pregnancy *Ann N Y Acad Sci* 148 501-531, 1968

GUILLEMIN, R, AND W VALE Bioassays of the hypophysiotropic hormones *in vitro* systems In *Hypophysiotropic Hormones of the Hypothalamus Assay and Chemistry*, edited by J Meites Baltimore Williams & Wilkins, 1970, p 21-35

HAGGENDAL, J An improved method for fluorimetric determination of small amounts of adrenaline and noradrenaline in plasma and tissues *Acta Physiol Scand* 59 242-254, 1963

HALMI, N S The accumulation and recirculation of iodide by the thyroid In *The Thyroid Gland*, edited by R Pitt Rivers and W R Trotter London Butterworths, 1964 vol 1, p 71-86

HAMPSON, J L, AND J G HAMPSON The ontogenesis of sexual behavior in man In *Sex and Internal Secretions*, 2, edited by W C Young Baltimore Williams & Wilkins, 1961 p 1401-1451

HARDING, B W, AND D H NELSON Electron carriers of the bovine adrenal cortical respiratory chain and hydroxylating pathways *J Biol Chem* 241 2212-2219 1966

HARRIS, G W Unsolved problems in the portal vessel chemotransmitter hypothesis In *Hypophysiotropic Hormones of the Hypothal*

amus Assay and Chemistry edited by J Meites Baltimore Wil liams & Wilkins 1970, p 1-20

HARRIS G W AND D JACOBSOHN Functional grafts of anterior pituitary gland *Proc Roy Soc (London) [Biol] Ser B* 139 263-276 1952

HECHTER O M M SOLOMON A ZAFFARONI, AND G PINCUS Transformation of cholesterol and acetate to adrenal cor tical hormones *Arch Biochem* 46 201-214 1953

HELLER C G AND Y CLERMONT Kinetics of the germinal epi thelium in man *Recent Progr Hormone Res* 20 545-575 1964

HELLER, H The hormone content of the vertebrate hypothalamo neurohypophysial system *Brit Med Bull* 22 227-231 1966

HERTZ R *Endocrine Effects of Tumors of Trophoblastic origin* London Excerpta Medica International Congress Series 83 Pro ceedings 2nd International Congress Endocrinology 1964 p 1068-1071

HODGES D R, AND W H McSHAN Electrophoretic separation of hormones associated with secretory granules from rat anterior pitui tary glands *Acta Endocr (Kobenhavn)* 63 378-384, 1970

HOLDEN M, AND L B ADAMS Inhibitory effects of cortisone ace tate and hydrocortisone on growth of fibroblasts *Proc Soc Exp Biol Med* 95 364-368 1957

HOUSSAY, B A AND A BIASOTTI Pankreasdiabetes und Hypo-physe beim Hund *Pflugers Arch Ges Physiol* 277 664 1931

HULTUNEN J K, D STEINBERG AND S E MAYER Protein kinase activation and phosphorylation of purified hormone sensitive lipase *Biochem Biophys Res Commun* 41 1350-1356 1970

HYMER W C AND W H McSHAN Isolation of rat pituitary gran ules and the study of their biochemical properties and hormonal ac tivities *J Cell Biol* 17 67-86, 1963

JAILER, J W Virilism *Bull N Y Acad Med* 29 377-394 1953

JEANRENAUD, B, AND D HEPP (eds) *Adipose Tissue Regulation and Metabolic Function Hormone and Metabolic Research, Supple ment 2* New York Academic Press, 1970

JEANRENAUD, B AND A E RENOLD Studies on rat adipose tissue *in vitro* VII Effects of adrenal cortical hormones *J Biol Chem* 235 2217-2223 1960

JEFFERSON L S, J H EXTON R W BUTCHER E W SUTH ERLAND AND C R PARK Role of adenosine 3',5' monophosphate in the effects of insulin and antiinsulin serum on liver metabolism *J Biol Chem* 243 1031-1038 1968

JENSEN, E V T SUZUKU, T KAWASHIMA, W E STUMPF P W JUNGBLUT AND E R DeSOMBRE A two step mechanism for the interaction of estradiol with rat uterus *Proc Nat Acad Sci U S A* 59 632-638, 1968

JOHNSON, J A, W W MOORE, AND W E SEGAR. The relation ship between left atrial pressure and plasma ADH levels in the dog *Physiologist* 10 213 1967

JOST, A Problems of fetal endocrinology of the gonadal and hypo physeal hormones *Recent Progr Hormone Res* 8 379-418, 1953

JUNGAS, R L, AND E G BALL Studies on the metabolism of adi pose tissue XII The effects of insulin and epinephrine on free fatty acid and glycerol production in the presence and absence of glucose *Biochemistry* 2 383-388, 1963

KALU D N J CAFREY, AND G V FOSTER *In vivo* effect of calci tonin parathyroid extract, and low calcium diet on bone collagen formation In *Calcitonin 1969* New York Springer Verlag, 1970 p 161-167

KAMBERI, I A, H P G SCHNEIDER, AND S M McCANN Ac tion of dopamine to induce release of FSH releasing factor (FRF) from hypothalamic tissue *in vitro Endocrinology* 86 278-284, 1970

KAPLAN, S A, AND C S SHIMIZU Effects of cortisol on amino acids in skeletal muscle and plasma *Endocrinology* 72 267-272, 1963

KATSOYANNIS, P G, K FUKUDA A TOMETSKO, K SUZUKI AND M TILAK Insulin peptides X The synthesis of the B chain of insulin and its combination with natural or synthetic A chain to generate insulin activity *J Amer Chem Soc* 86 930-932, 1964

KATZEN, H M, F TIETZE, AND D STETTEN, JR Further studies on the properties of hepatic glutathione insulin transhydrogenase *J Biol Chem* 238 1011-1066 1963

KAUFMAN, S, AND S FRIEDMAN Dopamine β hydroxylase *Pharmacol Rev* 17 71-100, 1965

KENDALL, E C The isolation in crystalline form of the compound containing iodine which occurs in the thyroid *J A M A* 64 2042 1915

KIPNIS D M Growth hormone and insulin antagonism In *Nature and Treatment of Diabetes* edited by B S Leibel and G A Wren shall New York Excerpta Medica Foundation 1965 p 258-273

KIRSCHNER, N H SAGE, W J SMITH AND A G KIRSCHNER Release of catecholamines from bovine adrenal medulla *Pharmacol ogist* 8 123 1966

KLEEMAN C R J W CZACZKES AND R CUTLER Mechanisms of impaired water excretion in adrenal and pituitary insufficiency IV Antidiuretic hormone in primary and secondary adrenal insuffi ciency *J Clin Invest* 44 1641-1648 1964

KLEEMAN C R M H MAXWELL, AND R E ROCKNEY Mech anisms of impaired water excretion in adrenal and pituitary insuffi ciency I The role of altered glomerular filtration rate and solute excretion *J Clin Invest* 37 1799-1808 1958

KOBAYASHI, H AND M TOKUZO Fine structure of the median eminence and its functional significance In *Frontiers in Neuroen docrinology* edited by W F Ganong and L Martini New York Oxford Univ Press 1969, p 3-46

KOERING M J Cyclic changes in ovarian morphology during the menstrual cycle in Macaca mulatta *Amer J Anat* 126 73-101 1969

KOPIN I J The adrenergic synapse In *The Neurosciences*, edited by G C Quarton, T Melnechuk and F O Schmitt New York Rockefeller University Press 1967, p 427-432

KORENMAN S G D TULCHINSKY, AND L W EATON JR Radio ligand procedures for estrogen assay in normal and pregnancy plasma In *Steroid Assay by Protein Binding* Karolinska Symposia on Research Methods in Reproductive Endocrinology 2nd 1970 p 291-304

KORNER, A Regulation of the rate of synthesis of messenger ribonu cleic acid by growth hormone *Biochem J* 92 449-456 1964

KORNER, A Ribonucleic acid and hormonal control of protein syn thesis *Progr Biophys* 17 61-98 1967

KOSTYO, J L, AND A F REDMOND Role of protein synthesis in the inhibitory action of adrenal steroid hormones on amino acid transport by muscle *Endocrinology* 79 531-540, 1966

KRAICER, J, AND J M MILLIGAN Suppression of ACTH release from adenohypophysis by corticosterone an *in vitro* study *Endocri nology* 87 371-376 1970

KRAICER J J V MILLIGAN J C GOSBEE, R G CONRAD, AND C M BRANSON Potassium corticosterone and adrenocorti cotropic hormone release *in vitro Science* 164 426, 1969

KREBS E G, R B HUSTON, AND F L HENKELER Properties of phosphorylase kinase and its control in skeletal muscle *Advances Enzym Regulat* 6 243-255, 1968

KREBS, H Gluconeogenesis (The Croonian lecture 1963) *Proc Roy Soc (London) [Biol] Ser B* 159 545-564, 1964

KYLE, L H, R J MEYER, M SCHAAF AND E J WERDEIN The effects of compound A (11 hydrocorticosterone) and compound F (17 hydroxycorticosterone) on total body fat *J Clin Invest* 35 1045-1054 1956

LACY, P E Beta cell secretion—from the standpoint of a pathobiolo gist *Diabetes* 19 895-905 1970

LANGAN, T A Phosphorylation of histones *in vivo* under the control of cyclic AMP and hormones In *Advances in Biochemical Psycho pharmacology*, edited by P Greengard and E Costa Press New York Raven, 1970 vol 3, p 307-323

LANMAN, J T An interpretation of human foetal adrenal structure and function In *The human adrenal cortex*, edited by A R Cirroe T Symington, and J K Grant Baltimore Williams & Wilkins 1962, p 547-558

LARDY, H A Gluconeogenesis pathways and hormonal regulation *Harvey Lect* 60 261-278, 1966

LARNER, J, C VILLAR PALASI, N D GOLDBERG, J S BISHOP, F HUIJING, J I WENGER, H SASKO, AND N B BROWN Hormonal and non hormonal control of glycogen synthesis —control of transferase phosphatase and transferase I kinase In *Control of Glycogen Metabolism* edited by W J Whelan Universi-teforlaget, Oslo New York Academic Press, 1968, p 1-18

LEE, K, J R REEL, AND F T KENNEY Regulation of tyrosine α-ketoglutarate transaminase in rat liver IX Studies of the mecha-nisms of hormonal inductions in cultured hepatoma cells *J Biol Chem* 245 5806-5812, 1970

LEFKOWITZ R J J ROTH AND I PASTAN Radioreceptor assay of adrenocorticotropic hormone new approach to assay of polypeptide hormones in plasma *Science* 170 633-635 1970

LEVINE R Mechanisms of insulin secretion *New Eng J Med* 283 522-526 1970

LEVINE R J The adrenal medulla and catecholamines In *Duncan's Diseases of Metabolism* edited by P K Bondy Philadelphia Saunders 1969 p 886-903

LEVINE R J AND H D CONN Tyrosine metabolism in patients with liver disease *J Clin Invest* 46 2012-2020 1967

LEVITT M S SPECTOR A SJOERDSMA AND S UDENFRIEND Elucidation of the rate limiting step in norepinephrine biosynthesis in the perfused guinea pig heart *J Pharmacol Exp Ther* 148 1-8, 1965

LI C H The chemistry of human pituitary growth hormone In *Growth Hormone* edited by A Pecile and E E Muller Foundation Amsterdam Excerpta Medica Foundation, 1968, p 3-28

LIAO S AND S FANG Receptor proteins for androgens and the mode of action of androgens on gene transcription in ventral prostate *Vitamins Hormones (N Y)* 27 17-90 1969

LIDDLE, G W Cushing s syndrome In *The Adrenal Cortex*, edited by A B Eisenstein Boston Little Brown 1967 p 523-551

LIDDLE G W L E DUNCAN JR, AND F C BARTTER Dual mechanism regulating adrenocortical function in man *Amer J Med* 21 380-386 1956

LIDDLE, G W D ISLAND, AND C K MEADOR Normal and abnormal regulation of corticotropin secretion in man *Recent Progr Hormone Res* 18 125-166, 1962

LIDDLE G W W E NICHOLSON D P ISLAND D N ORTH K ABE AND S C LOWDER Clinical and laboratory studies of ectopic humoral syndromes *Recent Progr Hormone Res* 25 283-305 1969

LINN T C F H PETTIT AND L J REED α keto acid dehydrogenase complexes X Regulation of the activity of the pyruvate dehydrogenase complex from beef kidney mitochondria by phosphorylation and dephosphorylation *Proc Nat Acad Sci USA* 62 234-241 1969

LIPSCOMB H S AND D H NELSON A sensitive biologic assay for ACTH *Endocrinology* 71 13-23 1962

LIPSETT, M B C J MIGEON, M A KIRSCHNER AND C W BARDIN Physiologic basis of disorders of androgen metabolism *Ann Intern Med* 68 1327-1344 1968

LIPSETT M B H WILSON M A KIRSCHNER, S G KORENMAN L M FISHMAN, G A SARFATY AND C W BARDIN III Steroid hormones studies on Leydig cell physiology and pathology secretion and metabolism of testosterone *Recent Progr Hormone Res* 22 245-281 1966

LOFFLER, G, AND L WEISS Lipase activities in adipose tissue In *Adipose Tissue Regulation and Metabolic Function Hormone and Metabolic Research, Supplement 2* edited by B Jeanrenaud and D Hepp New York Academic Press 1970 p 32-36

LONG C N H B KATZIN AND E G FRY The adrenal cortex and carbohydrate metabolism *Endocrinology* 26 309-344 1940

LUDWIG K S AND M HOROWITZ Anovular corpus luteum in a woman taking an oral contraceptive *Obstet Gynec* 33 696-702 1969

LUFT, R, AND E CERASI Human growth hormone as a regulator of blood glucose concentration and as a diabetogenic substance *Diabetologia* 4 1-9 1968

LUSE S Fine structure of the adrenal cortex In *The Adrenal Cortex* edited by A B Eisenstein Boston Little, Brown 1967, p 1-59

MACDONALD P C, R P ROMBAUT AND P K SIITERI Plasma precursors of estrogen I Extent of conversion of plasma Δ^4 androstenedione to estrone in normal males and non pregnant, normal, castrate, and adrenalectomized females *J Clin Endocr* 27 1103-1111, 1967

MACGREGOR, A G Hypothyrodism In *The Thyroid Gland*, edited by R Pitt-Rivers and W R Trotter London Butterworths, 1964 vol 2 p 112-129

MADISON, L L, D MEBANE, R H UNGER, AND A LOCHNER The hypoglycemic action of ketones II Evidence for a stimulatory feedback of ketones on the pancreatic beta cells *J Clin Invest* 43 408-415, 1964

MAGNUS LEVY, A. Ueber den respiratorischen Gewechsel unter dem Einfluss der Thyreoiden sowie unter verschiedenen pathologischen zustanden *Klin Wschr* 32 650-653 1895

MALLETTE L E J H EXTON AND C R PARK Control of gluconeogenesis from amino acids in the perfused rat liver *J Biol Chem* 244 5713-5723 1969

MANN F C AND T B MAGATH Die Wirkung der totalalen Leber erturpation *Ergebn Physiol* 23 212-262 1924

MARKS V AND E SAMOLS Intestinal factors in the regulation of insulin secretion In *Advances in Metabolic Disorders* edited by R Levine and R Luft New York Academic Press 1970 vol 4, p 1-38

MARLISS E T T AOKI P FELIG T POZEFSKY, AND G F CAHILL JR Hormones and substrates in the regulation of gluconeogenesis in fasting man *Advances Enzym Regulat* 8 3-11 1970

MARTIN, J B AND J REICHLIN Thyrotropin secretion in rats after hypothalamic electrical stimulation or injection of synthetic TSH releasing factor *Science* 168 1366-1368 1970

MARTINI L F FRASCHINI AND M MOTTA Neural control of anterior pituitary functions *Recent Progr Hormone Res* 24 439-485, 1968

MATTINGLY D A simple fluorimetric method for the estimation of free 11 hydroxycorticoids in human plasma *J Clin Path* 15 374-379 1962

MATTOX V R AND M LEWBART The determination of aldosterone in urine *J Clin Endocr* 19 1151-1161 1959

MAYES D S FURUYAMA D C KEM AND C A NUGENT A radioimmunoassay for plasma aldosterone *J Clin Endocr* 30 682-685 1970

MAYHEW, D A P H WRIGHT, AND J ASHMORE Regulation of insulin secretion *Pharmacol Rev* 21 183-212 1969

MCCANN S M Bioassay of LRF In *Hypophysiotropic Hormones of the Hypothalamus Assay and Chemistry* edited by J Meites Baltimore Williams & Wilkins 1970 p 90-102

MCCANN S M, AND J C PORTER Hypothalamic pituitary stimulating and inhibiting hormones *Physiol Rev* 49 240-284 1969

MCCRACKEN, J A A UNO, J R GODING, Y ICHKAWA AND D T BAIRD The *in vivo* effects of sheep pituitary gonadotrophins on the secretion of steroids by the autotransplanted ovary of the ewe *J Endocr* 45 425-440 1969

MCINTYRE N C D HOLDSWORTH AND D S TURNER Intestinal factors in the control of insulin secretion *J Clin Endocr* 25 1317-1324 1965

MCLEAN, P AND M W GURNEY Effect of adrenalectomy and growth hormone on enzymes concerned with urea synthesis in rat liver *Biochem J* 87 96-104 1963

MEITES, J (ed) *Hypophysiotropic Hormones of the Hypothalamus Assay and Chemistry* Baltimore Williams & Wilkins, 1970

MEITES, J, AND C S NICOLL Adenohypophysis prolactin *Ann Rev Physiol* 28 57-88 1966

MELICK, R A, G D AURBACH AND J T POTTS JR Distribution and half life of ^{131}I labeled parathyroid hormone in rat *Endocrinology* 77 198-202, 1965

MERCIER BODARD C A ALFSEN, AND E E BAULIEU Sex steroid binding plasma protein (SBP) Karolinska Symposia on Research Methods in Reproductive Endocrinology *Acta Endocr (Kobenhavn)* 147 (Suppl) 204-224, 1970

MICHELAKIS A M, J H FOSTER, G W LIDDLE, R K RHAMY, O KUCHE AND R. D GORDON Measurement of renin in both renal veins Its use in diagnosis of renovascular hypertension *Arch Intern Med (Chicago)* 120 444-448, 1967

MILLIGAN, J V, AND J KRAICER Adenohypophysial transmembrane potentials polarity reversal by elevated external potassium ion concentration *Science* 167 182-183 1970

MITTWOCH, U Sex chromatin (review) *J Med Gen* 1 50-76 1964

MOORE, K L (ed) *The Sex Chromatin* Philadelphia Saunders 1966

MORGAN, H E L S JEFFERSON, E B WOLPERT, AND D E RANNELS Regulation of protein synthesis in heart muscle II Effects of amino acid levels and insulin on ribosomal aggregation *J Biol Chem* 246 2163-2170 1971

MORGAN, H E, AND A PARMEGIANI Regulation of glycogenolysis in muscle III Control of muscle glycogen phosphorylase activity *J Biol Chem* 239 2440-2445, 1964

MORGAN H E, D M REGEN M J HENDERSON, T K SAW

YER, AND C R PARK Regulation of glucose uptake in muscle VI Effects of hypophysectomy adrenalectomy growth hormone hydrocortisone and insulin on glucose transport and phosphorylation in the perfused rat heart *J Biol Chem* 236 2162–2168 1961

MORGAN H E D M REGEN AND C R PARK Identification of a mobile carrier mediated sugar transport system in muscle *J Biol Chem* 239 369–374 1964

MORGANS M E Hyperthyroidism In *The Thyroid Gland* edited by R Pitt Rivers and W R Trotter London Butterworths 1964 vol 2 p 151–170

MORTIMORE, G E Influence of insulin on the hepatic uptake and release of glucose and amino acids In *Handbook of Physiology* ed ited by D F Steiner and N Freinkel In press 1971

MORTIMORE, G E AND C E MONDON Inhibition by insulin of valine turnover in liver *J Biol Chem* 245 2375–2383 1970

MOTTA M F PIVA F FRASCHINI AND L MARTINI Pituitary depletion methods for the bioassay of hypothalamic releasing factors In *Hypophysiotropic Hormones of the Hypothalamus Assay and Chemistry* edited by J Meites Baltimore Williams & Wilkins 1970 p 44–59

MULROW P J Aldosterone in hypertension and edema In *Duncan's Diseases of Metabolism*, edited by P K Bondy Philadelphia Saunders 1969 p 1083–1102

MUNRO H N Free amino acid pools and their role in regulation In *Mammalian Protein Metabolism* edited by H N Munro New York Academic Press, 1970, vol 4, p 299–386

MUNSON P L AND T K GRAY Function of thyrocalcitonin in normal physiology *Fed Proc* 29 1206–1208, 1970

MUNSON P L P F HIRSCH H B BREWER R A REISFELD, C W COOPER, A B WASTHED H ORIMO AND J T POTTS, JR Thyrocalcitonin *Recent Progr Hormone Res* 24 589–637 1968

MURPHY B E P Protein binding and the assay of nonantigenic hormones *Recent Progr Hormone Res* 25 563–610 1969

MURRAY I P C, AND E M McGIRR Iodine metabolism in thyroid dysfunction In *The Thyroid Gland* edited by R Pitt Rivers and W R Trotter London Butterworths 1964, vol 2, p 39–70

NAKANE P K Classifications of anterior pituitary cell types with immunoenzyme histochemistry *J Histochem Cytochem* 18 9–20, 1970

NALBANDOV A V Comparative aspects of corpus luteum function *Biol Reprod* 2 7–13 1970

NEELY J R R H BOWMAN, AND H E MORGAN Conservation of glycogen in the perfused rat heart developing intraventricular pressure pp 49–64 In *Control of Glycogen Metabolism*, edited by W J Whelan Universiteforlaget Oslo New York Academic Press 1968 p 49–64

NEELY J R, R H BOWMAN, AND H E MORGAN Effects of ventricular pressure development and palmitate on glucose transport *Amer J Physiol* 216 804–811 1969

NELSON, D H Eosinophiles, lymphocytes and 17 hydroxycorticosteroids as indices of adrenocortical activity In Ciba Foundation Colloquia Endocrinology V *Bioassay of Anterior Pituitary and Adrenocortical Hormones* Boston Little Brown, 1953, p 162–170

NEWSHOLME, E A, AND W GEVERS Control of glycolysis and gluconeogenesis in liver and kidney cortex In *Vitamins and Hormones*, edited by R S Harris, I G Wool, and J A Loraine New York Academic Press 1967 vol 25, p 1–87

NEY R L, E OGATA, N SHIMIZU, W E NICHOLSON, AND G W LIDDLE Structure function relationships of ACTH and MSH analogues In *Proceedings of the 2nd International Congress, Part 2* pp 1184–1191 Amsterdam Excerpta Medica Foundation 1965 p 1184–1191

NEY, R L D N ORTH, AND G W LIDDLE Evaluation of pituitary adrenal function in man *Mem Soc Endocr* 17 285–300, 1968

NOLAN, C, E MARGOLIASH, J D PETERSON AND J F STEINER The structure of bovine proinsulin *J Biol Chem* 246 2780–2795 1971

NORYMBERSKI, J K Methods of group corticosteroid estimation in urine II In *The Adrenal Cortex*, edited by G K McGowan and M Sandler Philadelphia Lippincott, 1961, p 90

NORYMBERSKI, J K, R D STUBBS, AND H F WEST Assessment of adrenocortical activity by assay of 17-ketogenic steroids in urine *Lancet* 1 1276–1281, 1953

ODELL W D J F WILBER, AND R. D UTIGER Studies of thyrotropin physiology by means of radioimmunoassay *Recent Progr Hormone Res* 23 47 78 1967

OJEDA S R, AND V D RAMIREZ Automatic control of LH and FSH secretion by short feedback circuits in immature rats *Endocrinology* 84 786–797 1969

O MALLEY B W Hormonal regulation of nucleic acid and protein synthesis *Trans N Y Acad Sci* 31 478–503 1969

O MALLEY B W Mechanisms of action of steroid hormones *New Eng J Med* 1971 In press

OMDAHL J M HOLICK T SUDA AND H F DELUCA Biological activity of 1 25 dihydroxycholecalciferol *Biochemistry* 10 2935–2940 1971

ORTH D N AND D P ISLAND Light synchronization of the circadian rhythm in plasma cortisol (17 OHCS) concentration in man *J Clin Endocr* 29 479–486 1969

ORTH D N D P ISLAND, W E NICHOLSON K ABE AND J P WOODHAM ACTH radioimmunoassay interpretation comparison with bioassay and clinical application In Radioisotopes in medicine *in vitro* studies edited by R L Hayes, F A Goswitz and B E P Murphy U S Atomic Energy Commission Division Technical Information 1968 p 251–272

OWEN M AND P J BINGHAM The effect of parathyroid extract on RNA synthesis in osteogenic cells *in vivo* In *Parathyroid Hormone and Thyrocalcitonin* edited by R V Talmage and L F Belanger New York Excerpta Medica Foundation 1968 p 216–225

OWEN O E, A P MORGAN H G KEMP J M SULLIVAN M G HERRERA AND G F CAHILL JR Brain metabolism during fasting *J Clin Invest* 46 1589–1595 1967

PAGE L B AND R B COPELAND Pheochromocytoma *DM* 1 1–40 1968

PALMORE W P AND P J MULROW Control of aldosterone secretion by the pituitary gland *Science* 158 1482–1484 1967

PARK C R, O B CROFFORD, AND T KONO Mediated (nonactive) transport of glucose in mammalian cells and its regulation *J Gen Physiol* 52 296s–318s 1968

PARK C R H E MORGAN M J HENDERSON, D M REGEN E CADENAS AND R L POST The regulation of glucose uptake in muscle as studied in the perfused rat heart *Recent Progr in Hormone Res* 27 493–529 1961

PARKER M L, AND W H DAUGHADAY Growth retardation correlating of plasma GH responses to insulin and arginine with subsequent metabolic and skeletal responses to GH treatment In *Growth Hormone* edited by A Pecile and E E Muller New York Excerpta Medica Foundation 1968 p 398–407

PARKER, M L, J M HAMMOND AND W H DAUGHADAY The arginine provocative test an aid in diagnosis of hyposomatotropism *J Clin Endocr* 27 1129–1136 1967

PASSONNEAU, J V, AND O H LOWRY Phosphofructokinase and the Pasteur effect *Biochem Biophys Res Commun* 7 10–15, 1962

PASSONNEAU, J V AND O H LOWRY P fructokinase and the control of the citric acid cycle *Biochem Biophys Res Commun* 13 372–379, 1963

PAULSEN, C A The testes In *Textbook of Endocrinology*, edited by R H Williams Philadelphia Saunders 1968, p 405–458

PEARSE, A G E The characteristics of the C cell and their significance in relation to those of other endocrine polypeptide cells and to the synthesis, storage and secretion of calcitonin In *Calcitonin 1969* New York Springer-Verlag, 1970, p 125–140

PENNY, D P, D I PATT AND W C DIXON, JR The fine structure of regenerating adrenocortical autotransplants in the rat *Anat Rec* 146 319–325, 1963

PERMUTT, M A C W PARKER, AND R D UTIGER Immunochemical studies with lysine vasopressin *Endocrinology* 78 809–814, 1966

PETERSON E A, AND W EVANS Separation of bone marrow cells by sedimentation at unit gravity Nature (*London*) 214 824–825, 1967

PETERSON, R. E The miscible pool and turnover rate of adrenocortical steroids in man *Recent Progr Hormone Res* 15 231–261, 1959

PITT RIVERS, R AND R. R CAVALIERI Thyroid hormone biosynthesis In *The Thyroid Gland*, edited by R Pitt Rivers and W R Trotter London Butterworths, 1964 vol 1 p 87–112

PORTER, C C, AND R H SILBER. A quantitative color reaction for

cortisone and related 17 21 dihydroxy 20 ketosteroids *J Biol Chem* 185 201-207 1950

PORTER J C B D GOLDMAN AND J F WILBER Hypophysiotropic hormones in portal vessel blood In *Hypophysiotropic Hormones of the Hypothalamus Assay and Chemistry* edited by J Meites Baltimore Williams & Wilkins 1970 p 282-297

POTTS J T JR Recent advances in thyrocalcitonin research *Fed Proc* 29 1200-1205 1970

POTTS J T, JR H T KEUTMANN H NIALL, I J DEFTOS, H B BREWER JR AND G D AURBACH Covalent structure of bovine parathyroid hormone in relation to biological and immunological activity In *Parathyroid Hormone and Thyrocalcitonin* edited by R V Talmage and L F Belanger New York Excerpta Medica Foundation 1968 p 44-53

POTTS, J T JR, H D NIALL H T KEUTMANN L J DEFTOS AND J A PARSONS Calcitonin recent chemical and immunological studies In *Calcitonin 1969* New York Springer Verlag 1970 p 56-73

POZEFSKY, T P FELIG J D TOBIN J S SOELDNER AND G F CAHILL JR Amino acid balance across tissues of the forearm in postabsorptive man Effects of insulin at two dose levels *J Clin Invest* 48 2273-2282 1969

PURVES H D Control of thyroid function In *The Thyroid Gland* edited by R Pitt Rivers and W R Trotter London Butterworths 1964 vol 2 p 1-38

RABINOWITZ, D The adrenal cortex In *The Principles and Practice of Medicine* edited by A M Harvey, L E Cluff R J Johnes A H Owens, Jr D Rabinowitz and R S Ross New York Appleton 1968 p 1058-1084

RACADOT J L Olivier E Porcile AND B DROZ Appareil de golgi et origine des grains de secretion dans les cellules adenohypophysaires chez le rat etude radiographique en microscopie electronique apus injection de leucine tritee *C R Acad Sci [D] (Paris)* 261 2972-2980, 1965

RALL J E J ROBBINS AND C G LEWALLEN The thyroid In *The Hormones*, edited by G Pincus K V Thimann and E B Astwood New York Academic Press 1964 p 159-439

RANDLE, P J, S J H ASHCROFT, AND J R GILL Carbohydrate metabolism and release of hormones In *Carbohydrate Metabolism and its Disorders*, edited by F Dickens, P J Randle, and W J Whelan London Academic Press 1968, p 427-447

RANDLE, P J P B GARLAND C N HALES E A NEWSHOLME R M DENTON, AND C I POGSON Interactions of metabolism and the physiological role of insulin *Recent Progr Hor mone Res* 22 1-48, 1966

RANNELS, D E, L S JEFFERSON, A C HJALMARSON E B WOLPERT, AND H E MORGAN Maintenance of protein synthesis in hearts of diabetic animals *Biochem Biophys Res Commun* 40 1110-1116 1970

RASMUSSEN, H, C ARNAUD AND C HAWKER Actinomycin D and the response to parathyroid hormone *Science* 144 1019-1021 1964

RASMUSSEN, H, AND A TENENHOUSE Cyclic adenosine monophosphate Ca^{++} and membranes *Proc Nat Acad Sci U S A* 59 1364-1370 1968

REID L J Pyruvate dehydrogenase complex In *Current Topics in Cellular Regulation*, edited by B L Horecker and E R Stadtman New York Academic Press 1969, vol 1, p 233-251

RENOLD, A E, AND G F CAHILL JR (eds) *Adipose tissue Handbook of Physiology* Washington, D C Amer Physiol Soc, 1965 sect 5

RESHEF L, R W HANSON, AND F J Ballard A possible physiological role of glyceroneogenesis in rat adipose tissue *J Biol Chem* 245 5979-5984, 1970

RESKO J A, AND K B EIK-NES Diurnal testosterone levels in peripheral plasma of human male subjects *J Clin Endocr* 26 573-576, 1966

RIGGS, T R Hormones and the transport of nutrients across cell membranes In *Actions of Hormones in Molecular Processes*, edited by G Litwack and D Kritchevsky New York Wiley 1964, p 1-57

ROBERTS, S The influence of adrenal cortex on the mobilization of tissue protein *J Biol Chem* 200 77-88, 1953

ROBBINS, J, AND J E RALL Proteins associated with the thyroid hormones *Physiol Rev* 40 415-489, 1960

ROBINSON D S AND D R WING Regulation of adipose tissue clearing factor lipase activity In *Adipose Tissue Regulation and Metabolic Function Hormone and Metabolic Research Supplement* 2 edited by B Jeanrenaud and D Hepp New York Academic Press 1970 p 41-46

ROBISON G A R W BUTCHER I OYE H E MORGAN AND E W SUTHERLAND The effect of epinephrine on adenosine 3 5 phosphate levels in the isolated perfused rat heart *Molec Pharmacol* 1 168-177 1965

ROBISON G A R W BUTCHER AND E W SUTHERLAND Cyclic AMP *Ann Rev Biochem* 37 149-174 1968

RODBARD D Mechanics of ovulation *J Clin Endocr* 28 849-861 1968

ROSEN F H R HARDING R J MILHOLLAND AND C A NICHOL Glucocorticoids and transaminase activity VI Comparison of the adaptive increases of alanine and tyrosine α ketoglutarate transaminase *J Biol Chem* 238 3726-3729 1963

ROSS B D R HEMS, AND H A KREBS The rate of gluconeogenesis from various precursors in the perfused rat liver *Biochem J* 102 942-951 1967

ROSS G T C M CARGILLE, M D LIPSETT, P L RAYFORD J R MARSHALL C A STROTT AND D RODBARD Pituitary and gonadal hormones in women during spontaneous and induced ovulatory cycles *Recent Progr Hormone Res* 26 1-62 1970

ROWSON L E A The evidence for luteolysin *Brit Med Bull* 26 14-16 1970

RUTHERFORD, M S AND D H NELSON Determination of urinary 17 ketogenic steroids by means of sodium metaperiodate oxidation *J Clin Endocr* 23 533-538, 1963

RYDIN H AND E B VERNEY The inhibition of water diuresis by emotional stress and muscular exercise *Quart J Exp Physiol* 27 343-374 1938

RYLE A P, F SANGER L F SMITH AND R KITAI The disulfide bonds of insulin *Biochem J* 60 541-556 1955

SAFFRAN M, AND A V SCHALLY The release of corticotropin by anterior pituitary tissue *in vitro Canad J Biochem* 33 408-415 1955

SALMI M, AND I I GESCHWIND Some effects of the hypothalamic luteinizing hormone releasing factor on the biosynthesis and release of luteinizing hormone *Endocrinology* 81 835-848, 1967

SAMUELS, L T AND T UCHIKAWA Biosynthesis of adrenal steroids In *The Adrenal Cortex*, edited by A B Eisenstein Boston Little Brown, 1967, p 61-102

SANDERSON, P H, F MARSHALL II AND R E WILSON Calcium and phosphorus homeostasis in the parathyroidectomized dog evaluation by means of ethylenediamine tetraacetate and calcium tolerance tests *J Clin Invest* 39 662-670 1960

SANGER, F Chemistry of insulin *Science* 129 1340-1344, 1959

SATO G Uber die Beziehungen des Diabetes insipidus zum Hypophysenhinterlappen und zum Tuber cinereum *Arch Exp Path Pharmak* 131 45-69 1928

SAYERS, G The adrenal cortex and homeostasis *Physiol Rev* 30 241-320 1950

SCHALLY, A V, A ARIMURA, A KASTIN, H MATSUO Y BABA, T REDDING, R NAIR, L DEBELJUK, AND W WHITE Gonadotropin releasing hormone One polypeptide regulates secretion of luteinizing and follicle stimulating hormones *Science* 173 1036-1038 1971

SCHALLY, A V, A ARIMURA, C Y BOWERS, A J KASTIN, S SAWANO AND T W REDDING Hypothalamic neurohormones regulating anterior pituitary function *Recent Progr Hormone Res* 24 497-581 1968

SCHALLY A V, A ARIMURA, C Y BOWERS, I WAKABASHI, T W REDDING, J C MITTLER P PIZZOLATO, AND A J SEPGAL Characterization of releasing hormones from human hypothalami (Program of the Endocrine Society) *Endocrinology* 86 1473, 1970a

SCHALLY A V A ARIMURA, I WAKABAYASHI, S SAWANO, J F BARRETT C Y BOWERS, T W REDDING, J C MITTLER, AND M SAITO The chemistry of hypothalamic growth hormone releasing hormone In *Hypophysiotropic Hormones of the Hypothalamus Assay and Chemistry*, edited by J Meites Baltimore Williams & Wilkins, 1970 p 208-225

SCHIMKE, R T Regulation of protein degradation in mammalian

tissues In *Mammalian Protein Metabolism* edited by H N Munro New York Academic Press 1970 vol 4 p 177-228

SCHIMKE R T E W SWEENEY AND C M BERLIN An analysis of the kinetics of rat liver tryptophan pyrrolase induction the significance of both enzyme synthesis and degradation *Biochem Biophys Res Commun* 15 214-219 1964

SCHNEIDER H P G, AND S M McCANN Release of LH releasing factor (LRF) into the peripheral circulation of hypophysectomized rats by dopamine and its blockage by estradiol *Endocrinology* 87 249-253 1970

SCHOENHEIMER R *The Dynamic State of Body Constituents* Harvard Univ Press Cambridge Mass 1942

SCHOENHEIMER, R AND D RITTENBERG The study of intermediary metabolism of animals with the aid of isotopes *Physiol Rev* 20 218-248 1940

SCHONEMANN A Hypophysis und Thyreoida *Arch Anat Path* 129 310-336 1892

SCRUTTON M C AND M F UTTER The regulation of glycolysis and gluconeogenesis in animal tissues *Ann Rev Biochem* 37 249-302 1968

SERRA G B, AND A R MIDGLEY The *in vitro* release of LH during continuous superfusion of single rat anterior pituitary glands *Proc Soc Exp Biol Med* 133 1370-1374 1970

SHAFRIR E, A GUTMANN E GORIN AND M OREVI Aspects of carbohydrate metabolism in adipose tissue In *Adipose Tissue Regulation and Metabolic Function Hormone and Metabolic Research Supplement 2* edited by B Jeanrenaud and D Hepp New York Academic Press 1970 p 130-135

SHALLABARGER, C J The effect of thyroid hormones on growth and differentiation In *The Thyroid Gland* edited by R Pitt Rivers and W R Trotter London Butterworths 1964 vol 1 p 187-198

SHARE L Effects of carotid occlusion and left atrial distension on plasma vasopressin titer *Amer J Physiol* 208 219-223, 1965

SHARE, L Extracellular fluid volume and vasopressin secretion In *Frontiers in Neuroendocrinology*, 1969 edited by W F Ganong and L Martini New York Oxford Univ Press 1969, p 183-210

SHARP, G W G, AND A LEAF Mechanism of action aldosterone *Physiol Rev* 46 593-633 1966

SHIMIZU K M HAYANO, M GUT AND R I DORFMAN The transformation of 20α hydroxycholesterol to isocaproic acid and C_{21} steroids *J Biol Chem* 236 695-699, 1961

SHIMIZU K S SHIMAO, AND M TANAKA Conversion *in vitro* of 20α hydroxycholesterol to 17α,20α dihydroxycholesterol by human fetal adrenals *Steroids* 1 (Suppl) 85-93 1965

SIEBER, P, M BRUGGER, B KAMBER, B RINIKER, W RITTEL, R. MAIER, AND M STAEHELIN Synthesis and biological activity of peptide sequences related to porcine α thyrocalcitonin In *Calcitonin 1969* New York Springer Verlag, 1970 p 28-33

SILBER, R H AND C C PORTER Nitrogen balance liver protein repletion and body composition of cortisone treated rats *Endocrinology* 52 518-525 1953

SLOPER, J C Hypothalamic neurosecretion *Brit Med Bull* 22 209-215 1966

SMITH, L F Species variation in the amino acid sequence of insulin *Amer J Med* 40 662-666, 1966

SMITH R E AND M G FARQUHAR Lysosome function in the regulation of the secretory process in cells of the anterior pituitary gland *J Cell Biol* 31 319-347, 1966

SMITH, R E, AND M G FARQUHAR Modulation in nucleoside diphosphatase activity of mammotrophic cells of the rat adenohypophysis during secretion *J Histochem Cytochem* 18 237-250 1970

STANBURY J B Familial goiter In *The Metabolic Basis of Inherited Disease*, edited by J B Stanbury, J B Wyngaarden, and D S Fredrickson New York McGraw Hill, 1966, p 273-320

STARLING, E H On the chemical correlation of the functions of the body *Lancet* 2 339-341, 1905

STEINBERG, D, AND M VAUGHN Release of free fatty acid from adipose tissue *in vitro* in relation to rates of triglyceride synthesis and degradation In *Adipose Tissue Handbook of Physiology* edited by A E Renold and G F Cahill Jr Washington, D C Amer Physiol Soc, 1965, sect 5, p 335-347

STEINBERG, J, AND G NICHOLS, JR Stimulation of bone RNA synthesis by parathyroid hormone In *Parathyroid hormone and thyrocalcitonin* edited by R V Talmage and L F Belanger New York Excerpta Medica Foundation 1968 p 226-241

STEINBERGER E Hormonal control of mammalian spermatogenesis *Physiol Rev* 51 1-22 1971

STEINER, D F J L CLARK C NOLAN A H RUBENSTEIN E MARGOLIASH B ATEN AND P E OYER Proinsulin and the biosynthesis of insulin *Recent Progr Hormone Res* 25 207-282 1969

STEINER D F AND J KING Induced synthesis of hepatic uridine diphosphate glucose—glycogen glucosyltransferase after administration of insulin to alloxan diabetic rats *J Biol Chem* 239 1292-1298, 1964

STERLING K Thyroxine in blood *Mayo Clin Proc* 36 586-608 1964

STERLING K The significance of circulating triiodothyronine *Recent Progr Hormone Res* 26 249-274 1970

STROTT C A T YOSHIMI G T ROSS AND M B LIPSETT Ovarian physiology relationship between plasma LH and steroidogenesis by the follicle and corpus luteum effect of HCG *J Clin Endocr* 29 1157-1167 1969

STRUCK E, J ASHMORE, AND O WIELAND Stimulierung der gluconeogenese durch langkettige fettsauren und glucagon *Biochem Z* 343 107-110, 1965

SUTHERLAND, E W On the biological role of cyclic AMP *JAMA* 214 1281-1288 1970

SUTHERLAND, E W I OYE, AND R W BUTCHER The action of epinephrine and the role of adenyl cyclase system in hormone action *Recent Progr Hormone Res* 21 623-646 1965

SUTHERLAND, E W, AND T W RALL The relation of adenosine 3 5 phosphate and phosphorylase to the actions of catecholamines and other hormones *Pharmacol Rev* 12 265-299, 1960

SWERDLOFF R S AND W M ODELL Serum luteinizing and follicle stimulation hormone levels during sequential and nonsequential contraceptive treatment of eugonadal women *J Clin Endocr* 29 157-163 1969

SZENT GYORGYI A J Observations on the function of the peroxidase system and the chemistry of the adrenal cortex description of new carbohydrate derivative *Biochem J* 22 1387-1409 1928

TANAKA Y H F DELUCA J OMDAHL AND M F HOLICK Mechanism of action of 1 25 dihydroxycholecalciferol on intestinal calcium transport *Proc Nat Acad Sci* 68 1286-1288, 1971

TATA J R Distribution and metabolism of thyroid hormones In *The Thyroid Gland* edited by R Pitt-Rivers and W R Trotter London Butterworths 1964, vol 1, p 163-186

TAUROG A Thyroid peroxidase and thyroxine biosynthesis *Recent Progr Hormone Res* 26 189-241 1970

TEMPLE, T E AND G W LIDDLE Inhibition of adrenal steroid biosynthesis *Amer Rev Pharmacol* 10 199-218, 1970

TEPPERMAN, J, AND H M TEPPERMAN Gluconeogenesis lipogenesis and the Sherringtonian metaphor *Fed Proc* 29 1284-1293, 1970

TOBIAN, L Interrelationship of electrolytes juxtaglomerular cells and hypertension *Physiol Rev* 40 280-312, 1960

TOMIZAWA H H Mode of action of an insulin degrading enzyme from beef liver *J Biol Chem* 237 428-431 1962

TOPPER, Y J Multiple hormone interactions in the development of mammary gland *in vitro* *Recent Progr Hormone Res* 26 287-303, 1970

TRENDELENBURG P Anteil der Hypophyse und des Hypothalamus am experimentellen Diabetes Insipidus *Klin Wschr* 7 1679-1680 1928

TURTLE, J R AND D M KIPNIS An adrenergic receptor mechanism for the control of cyclic 3 ,5 adenosine monophosphate synthesis in tissues *Biochem Biophys Res Commun* 28 797-802, 1967

UNDENFRIEND, S *Physiological Regulation of Noradrenaline Biosynthesis Adrenergic Neurotransmission* Boston Ciba Foundation Study Group, 1968, vol 33, p 3-11

VAES, G Parathyroid hormone like action of N^6-2'-O Dibutyryladenosine 3',5' (cyclic) monophosphate on bone explants in tissue culture *Nature (London)* 219 939-940, 1968a

VAES, G The role of lysosomes and of their enzymes in the development of bone resorption induced by parathyroid hormone In *Parathyroid Hormone and Thyrocalcitonin*, edited by R V Talmage and

L F Belanger New York Excerpta Medica Foundation 1968b p 318-328

VALE W AND R GUILLEMIN Potassium induced stimulation of thyrotropin release *in vitro* Requirement for presence of calcium and inhibition by thyroxine *Experentia* 23 855-857 1967

VALLATTON M B L B PAGE AND E HABER Radioimmunoassay of angiotensin in human plasma *Nature (London)* 215 714-715 1967

VANDER A J AND R MILLER Control of renin secretion in the anesthetized dog *Amer J Physiol* 207 537-546 1964

VANDEWIELE R L J BOGUMIL I DYRENFURTH M FERIN R JEWELEWICZ M WARREN T RIZKALLAH, AND G MIKHAIL Mechanisms regulating the menstrual cycle in women *Recent Progr Hormone Res* 26 63-103 1970

VAN WYK J J AND M M GRUMBACH Disorders of sex differentiation In *Textbook of Endocrinology* edited by R H Williams Philadelphia Saunders 1968 p 537-612

VARMUS H E R L PERLMAN AND I PASTAN Regulation of lac messenger ribonucleic acid synthesis by cyclic adenosine 3 5 monophosphate and glucose *J Biol Chem* 245 2259-2267, 1970

VERMEULEN A L VERDONCK, M VAN DER STRAETEN AND N ORIE Capacity of the testosterone binding globulin in human plasma and influence of specific binding of testosterone on its metabolic clearance rate *J Clin Endocr* 29 1470-1480 1969

VERNEY, E B The antidiuretic hormone and the factors which determine its release *Proc Roy Soc (London)* [*Biol*] *Ser B* 135 25-105 1947

VERNIKOS DANELLIS J AND B H MARKS The assay of corticotropin releasing factor In *Hypophysiotropic Hormones of the Hypothalamus Assay and Chemistry*, edited by J Meites Baltimore Williams & Wilkins 1970 p 60-73

VILLAR PALASI, C AND S H WEI Conversion of glycogen phosphorylase b to a by non activated phosphorylase b kinase an *in vitro* model of the mechanism of increase in phosphorylase a activity with muscle contraction *Proc Nat Acad Sci U S A* 67 345-350 1970

VILLEE, D B Development of endocrine function in the human placenta and fetus *New Eng J Med* 281 473-484 and 533-541, 1969

VON MERRING J V AND O MINKOWSKI *Arch Exp Path Pharmak* 26 375 1889-1890

WALKER, B E, AND E G RENNELS Adrenal cortical cell replacement in the mouse *Endocrinology* 68 365-374, 1961

WALSER, M Ion association VI Interactions between calcium magnesium, inorganic phosphate citrate, and protein in normal human plasma *J Clin Invest* 40 723-730, 1961

WASSERMAN R H AND A N TAYLOR Vitamin D_3-induced calcium binding protein in chick intestinal mucosa *Science* 152 791-793, 1966

WEBER, G Study and evaluation of regulation of enzyme activity and synthesis in mammalian liver *Advances Enzym Regulat* 1 1-35 1963

WEBBIN H AND I L CHAIKOFF Utilization of adrenal gland cholesterol for synthesis of cortisol by the intact normal and ACTH treated guinea pig *Arch Biochem* 93 476-482, 1961

WETTSTEIN, A., F W KAHNT, AND R. NEHER *The Biosynthesis of Aldosterone (Electrocortin) in the Adrenal* Boston Ciba Foundation Colloquia Endocrinology, 1955 vol 8, p 170

WICKS, W D Induction of hepatic enzymes by adenosine 3' 5' monophosphate in organ culture *J Biol Chem* 244 3941-3950 1969

WIELAND, O Ketogenesis and its regulation *Advances Metab Dis* 3 1-47 1968

WIELAND O AND E SIESS Interconversion of phospho and di phosphoforms of pig heart pyruvate dehydrogenase *Proc Nat Acad Sci U S A* 65 947-954 1970

WILBER, J F Stimulation of thyrotropin (TSH) synthesis by thyrotropin releasing factor (TRF) and 59 mM K^+ (Program of Endocrine Society) *Endocrinology* 86 1188, 1970

WILLIAMS R H The pancreas In *Textbook of Endocrinology* Ed 4 edited by R H Williams Philadelphia Saunders 1968 p 613-802

WILLIAMSON J R Effects of fatty acids glucagon and anti insulin serum on the control of gluconeogenesis and ketogenesis in rat liver *Advances Enzym Regulat* 5 229-255, 1967

WILLIAMSON J R E T BROWNING AND R SCHOLZ Control mechanisms of gluconeogenesis and ketogenesis I Effects of oleate on gluconeogenesis in perfused rat liver *J Biol Chem* 244 4607-4616 1969a

WILLIAMSON J R, R A KREISBERG AND P W FELTS Mechanism for the stimulation of gluconeogenesis by fatty acids in perfused rat liver *Proc Nat Acad Sci U S A* 56 247-254 1966

WILLIAMSON J R, R SCHOLZ AND E T BROWNING Control mechanisms of gluconeogenesis and ketogenesis II Interactions between fatty acid oxidation and the citric acid cycle in perfused rat liver *J Biol Chem* 244 4617-4627 1969b

WITSCHI, E Sex chromatin and sex differentiation in human embryos *Science* 126 1288-1290 1957

WOLFF, E C, AND J WOLFF The mechanism of action of the thyroid hormones In *The Thyroid Gland* edited by R Pitt Rivers and W R Trotter London Butterworths 1964 vol 1, p 237-282

WOLFF J Transport of iodide and other anions in the thyroid gland *Physiol Rev* 44 45-90 1964

WOLLMAN S H Secretion of thyroid hormones In *Lysosomes in Biology and Pathology*, edited by J T Dingle and H B Fell New York Elsevier, 1969, vol 2 p 483-512

WOODBURY D M Relation between the adrenal cortex and the central nervous system *Pharmacol Rev* 10 275-357 1958

WOODSIDE K H, AND G E MORTIMORE Control of proteolysis in the perfused rat liver influence of amino acids insulin and glucagon *Fed Proc* 29 379, 1970

WOOL, I G, AND E I WEINSHELBAUM Incorporation of C^{14} amino acids into protein of isolated diaphragms role of the adrenal steroids *Amer J Physiol* 197 1089-1092, 1959

WUNSCH, E Die totalsynthese des pankreas hormones glucagon *Z Naturforsch* [*B*] 22b 1269-1276 1967

WURTMAN R J Catecholamines *New Eng J Med* 273 637-646 693-700 746-753 1965

WURTMAN, R J AND J AXELROD Control of enzymatic synthesis of adrenaline in the adrenal medulla by adrenal cortical steroids *J Biol Chem* 241 2301-2305, 1966

YALOW, R S, AND S A BERSON Radioimmunoassay of human growth hormone in plasma In *Growth Hormone* edited by A Pecile and E E Muller New York Excerpta Medica Foundation 1968, p 60-83

YALOW R S, S M GLICK, J ROTH AND S A BERSON Radioimmunoassay of human plasma ACTH *J Clin Endocr* 24 1219-1225, 1964

YOSHIMI T, C A STROTT J R MARSHALL, AND M B LIPSETT Corpus luteum function in early pregnancy *J Clin Endocr* 29 225-230, 1969

ZIMMERMAN W Eine Farbreaktion der Sexualhormone und ihre Anwerdung zur quantitativen Colorimetrischen Bestimmung *Hoppe Seyler Z Physiol Chem* 233 257-264, 1935

8

Sensory Systems

1 / SENSORY PROCESSES

The adaptation of man to his environment and to the society of his fellow men requires the constant processing of information received from the environment and from individuals within the environment Any form of activity constitutes an interaction process wherein any kind of movement or action produces change The detection of such a change via the sensory systems permits the regulation and control of subsequent actions Much of consciousness, even in the absence of significant sensory inputs, seems to involve sensory imagery We think in terms of visual and auditory imagery and imagery involving the other senses Although in the analytic study of sensory processes there is often a tendency to consider them as independent dimensions, our perception of the environment and our relations to it depend in fact upon many complex interactions of the individual sensory modalities (Gibson, 1966) It is our purpose in the material which follows to consider the various important sensory processes individually for the purpose of understanding their nature, acknowledging the fact that they do not function independently but provide a collective basis for consciousness and the control of our activities

THE RECEPTOR ORGANS

The different sensory modalities are distinguished by specialized receptors which mediate responses to stimuli appropriate for the transfer of information in each of the individual sensory domains Sensory receptors are sometimes divided among each of three classifications, the *exteroceptors*, the *interoceptors*, and the *proprioceptors* The *exteroceptors* respond to stimuli that arise outside of the body, such as the light which stimulates the eye, sound pressure which stimulates the ear, various forms of mechanical pressure or contact with objects of a temperature different from that of the skin which stimulate the receptors of touch, minute quantities of chemical substances which stimulate the olfactory sense organs, and the receptor organs of taste The *interoceptors* lie within the mucous linings of the respiratory or digestive tract and respond to materials ingested or inhaled and to changes in pressure The *proprioceptors* are sensory receptor endings that respond to stimuli generated by muscular movement of the body or changes in muscular tension They reside in skeletal muscle, in tendons, in joints, in the heart, the carotid sinus, and in the gastrointestinal wall The vestibular receptors of the nonauditory labyrinth respond to linear and angular accelerations, that may be imposed upon the passive organism or may be generated by motor acts of the organism These receptors work in close harmony with the proprioceptors of muscle and are sometimes included with this category They are also extremely important in the control of eye movements, however, and hence for visual perception

THE STIMULUS

Classically, the physical form of stimulation that is most natural or appropriate for a given sensory receptor has been called the *adequate stimulus* This terminology has resulted in the use of the term *inadequate stimulus* for many forms of physical stimulation which may excite a given receptor system, but which are not likely to provide stimulation under normal circumstances In this sense then, inadequate stimuli are perfectly adequate to produce a response and the terminology is somewhat confusing In the material which follows we will refer to the *appropriate stimulus* rather than the *adequate stimulus* in order to avoid possible confusion

An extremely important aspect of the study of sensory processes is the definition of the nature of the appropriate stimulus The physical dimensions of the stimulus and the range of energies to which the receptor will respond, when an energetic continuum can be identi-

fied, must be determined in order to understand the nature of the function of peripheral receptor organs

CONSCIOUS CONCOMMITANTS OF SENSATION

The prominence of a sensory dimension in our consciousness varies significantly from one sensory modality to another Prominence can not be assigned any fixed order either, it may vary with circumstances In a normal, awake subject, with a well-lighted environment, the visual domain will normally be of great prominence Obviously, this will not be so if there is no light, if the subject's eyes are for some reason covered or closed, or if the subject is blind The auditory dimension may or may not be prominent, depending on the level of stimulation For someone listening with concentration to a musical work or engrossed in a conversation, the auditory dimension may be more prominent than the visual For someone engaged in any form of athletic competition which involves complex movements of the body, the vestibular sense will be playing an extremely critical role under normal conditions but will have no conscious concommitant which can be uniquely associated by the individual with that sensory dimension The same may not be true for organisms other than man in which vestibular function is more important These include flying birds, arboreal creatures and fish, animals with much greater mobility in three dimensions than man has Olfactory and gustatory senses may be prominent under certain circumstances, but their prominence tends to be relatively infrequent and associated with such actions as eating and exposure to unusual sources of stimulation Other classes of receptors that respond to chemical stimuli, those in the carotid sinus and the aortic bodies, afford an important basis for control of physiological processes but never intrude into consciousness Receptors for the sensation of pain, whether they reside in the skin or deeper within the bodily structure, can provide a compelling conscious component of sensation but, fortunately, under normal circumstances they are relatively inactive Their role is usually a protective one in providing information to assist the organism in removing itself from sources of potential injury

Although it is convenient to think of the senses categorically, it is most important to recognize that our perceptions of the physical world in which we live are dependent upon multisensory inputs and interactions among the senses Thus, the interpretation of visual stimuli may frequently be influenced by vestibular stimuli, proprioceptive stimuli and possibly auditory stimuli The developing organism learns to expect certain correlations among different stimulus dimensions As we shall see, certain unique situations which may result in a noncorrespondence of different stimuli with expected relationships may be extremely disconcerting or even debilitating Such noncorrespondences may be encountered between visual and vestibular inputs during exposure to the motions of a ship on a heavy sea or the complex flight path of a maneuvering aircraft Noncorrespondence between visual and tactual or muscle senses may arise during attempts to move about in a "distorted room" of the type one may find in an amusement park

METHODS OF STUDYING SENSORY SYSTEMS

Psychophysical Studies. The mode of functioning of sensory processes has fascinated man for hundreds of years and he has devised an ingenious variety of methods for their study For any situation where a human subject can detect stimulation of a given sense receptor, the function of the receptor may be studied with the intact human organism as an indicator of the effect of stimulation It is thus possible to explore the appropriate stimulus in terms of its physical dimensions along various continua In the case of vision, the subject's task may simply be to indicate when he is able to detect the presentation of a stimulus The investigator may then explore such variables as the wavelength of the stimulus, its physical intensity, the region of the retina on which the stimulus is projected, the size of the stimulus, and the temporal duration of its exposure Obviously, these various dimensions may be expected to interact and their interactions can also be studied The results of many such investigations of the visual sensory system are documented in the chapter on vision Similar investigations of the physical dimensions of the stimulus have been carried out for hearing,

touch, and the acceleration forces which stimulate the vestibular mechanism It is not possible to isolate physical continua associated with olfactory and gustatory stimulation, and so investigations of these senses have not been organized around the physical dimensions of the stimulus Rather, investigators have attempted to catalog and look for similarities among chemical stimuli which provide similar qualitative experiences in these sensory dimensions In addition to the minimum required physical energy, or concentration of a chemical substance, necessary for the threshold detection of stimulation, the investigator may wish to explore the effect of increasing stimulation along an intensive or concentration dimension and note the effect on the subject Limits of stimulation may be defined at which pain is experienced, or further increase in the appropriate dimension of the stimulus is accompanied by no further increase in the subjective intensity of the sensation Such studies may be carried out by having subjects adjust a matching stimulus to the experimental stimulus or by requiring them to make a subjective rating of stimulus intensity Philosophical objections may be raised in connection with the latter procedure, but various rating techniques have provided quite consistent results when properly used (Stevens and Galanter, 1957)

Electrophysiological Studies. Studies of sensory systems with human subjects possess certain important limitations It is difficult to decide what part of the overall sensory system may be determining the response to stimulation For example, is the maximum audible frequency which can be discriminated by man limited in the peripheral receptor for hearing or somewhere between the receptor and the cortical centers involved in hearing? It has been possible in man to record electrical activity arising directly in the retina as a result of photic stimulation Such activity, recorded as the electroretinogram, permits the limited study of purely peripheral processes in the visual system of man Unfortunately, similar indices of peripheral activity are not readily available for the other sensory systems

Sensory processes may be studied in animals with measurement of response at any desired point within the nervous system Techniques have been developed for the insertion into single cells of tiny electrodes with tip diameters even smaller than 1 μ when necessary In the visual system, in at least one organism, intracellular recordings have been obtained from each of the known cell types of the retina (Dowling, 1970) Electrodes need not be inserted into a cell in order to record the activity of a single cell Relatively small electrodes, with tip diameter of from 1 to 5 or more μ, located in the vicinity of a single cell or a single fiber may pick up signals which reflect the activity of that cell or fiber alone The higher the impedance of the electrode, the better is its discrimination With the higher impedance, however, it may be somewhat more difficult to make an initial detection of the activity of any cell

Sensory Coding. In recent years, studies of the activity of single cells and single fibers in the sensory nervous system have provided important new insights into the nature of sensory function It has become apparent that as sensory information is transmitted from the primary receptor through the various neural elements which link the receptor to cortical centers for the mediation of sensory response, there is a kind of selective convergence from certain receptors such that cells higher in the nervous system may respond selectively to unique characteristics of the stimulus dimension (Barlow, 1969) For example, there are cells in the cortex of higher animals which respond to a moving stimulus with a certain orientation in the visual world In hearing, specific cells may respond to certain patterns of change in the frequency or intensity of an auditory stimulus Others may respond maximally when a stimulus is presented to one ear shortly before it is presented to the other, as is the case for a sound source located on one side of the sensing organism Such cells selective to the sequence of lateral stimulation have even been reported for the olfactory system It thus appears that the nervous system is abstracting out certain dimensions of the stimulus world In some way the brain succeeds in putting these dimensions back together in a fashion that permits us to achieve veridical perceptions of the world in which we live

Animal Research Animals offer the advantage of permitting the study of the action of sensory systems at any location along the way from the primary receptor to the highest cortical centers identified for a given sensory system Such studies may involve microelec-

trodes as described above, or more gross electrodes that mirror the activity of relatively large populations of cells Both types of investigation are of value The procedure to be preferred depends upon the nature of the questions an experimenter wishes to ask

It may seem reasonable to study overall sensory experience in human subjects and to examine various portions of sensory systems electrophysiologically in animals The latter work may provide an understanding of why it is that certain results are obtained with the human subject for stimulation of a given sensory dimension There are serious objections however, to extrapolating from electrophysiological studies of animals directly to the intact human Certain electrical signals recorded in a sensory system may be merely incidental to stimulation, a kind of by-product of the significant aspects of the process, and hence not a direct linking process in the overall function of the sensory system The cochlear microphonic potential, discussed in some detail in the chapter on hearing, is probably such an incidental effect Another serious problem rests with the fact that the sensory processes of animals are not identical to those of the human In recent years, highly significant differences in the way in which visual information is processed in the sensory systems of fish, frogs, birds, rabbits, cats, and monkeys have been revealed

Presumably, the closer to man phylogenetically the experimental animal under study, then the more valid may be extrapolations from electrophysiological studies of that animal to the sensory processes of man Students of sensory systems however, have not been content with this kind of reasoning Another approach is to study the whole animal just as man himself is studied in sensory experiments The animal cannot verbalize his responses, but he can be trained to respond affirmatively or negatively if the experimenter provides him with an adequate reason to do so The reason may be a food reward or the avoidance of a shock or a few drops of water or some other appropriate *reinforcement* for a correct response Studies then can be conducted in which animals will signal appropriately if they detect a stimulus If, in behavioral studies of sensory processes with animals, the results are similar to, or identical with those found in man, it is more reasonable to assume that the underlying physiological processes studied by means of electrical recording techniques may be the same as those which exist in man

Even in animals that may be difficult to train to respond to stimulation, behavioral studies can be conducted This is accomplished by employing some natural reflex with which the animal is already equipped to respond to certain stimulus dimensions A favorite reflex for the study of visual processes has been the optokinetic reflex This is a tendency for an animal to turn in the direction of motion of a repeating pattern such as vertical stripe The head may turn, the whole body may turn, or the eyes may turn Obviously, if the stripes are not seen because they are too dim or too fine or because the contrast is too low, then the response will not appear

In behavioral studies, it is usual to train an animal to perform some special response when it makes a desired sensory discrimination Some dimension of the stimulus, such as the wavelength of light or the frequency of sound, may then be changed and the threshold for a discrimination redetermined under these new conditions The experimenter assumes that somewhere in the nervous system a criterion event of some sort has occurred which permits the animal to respond appropriately It is probably not true that events within the nervous system are identical for a constant threshold criterion when one alters the dimensions of the stimulus that are known to result in qualitative changes (Zacks, 1970) Nonetheless, this kind of assumption is frequently made and it is the basis for assuming that the results of experiments which explore systematically the influence of changing some stimulus dimension may be plotted as continuous functions when a constant criterion, such as detection threshold, is employed (Brindley, 1960)

In order to compare electrophysiological studies with behavioral studies of this sort, a fixed criterion is also adopted In the case of electrophysiological studies, the criterion may be a certain rate of discharge of impulses by a single cell somewhere in the nervous system The intensity of an auditory stimulus may be adjusted in order to achieve this rate of discharge for various frequencies of stimulation, or the intensity of the visual stimulus may be adjusted for each of a number of wavelengths in order to achieve this rate of discharge if the cell is one in the visual system In general,

results appear to be more nearly comparable with behavioral studies when such procedures are used in electrophysiological studies An alternative approach is to hold the energy of the stimulus constant while such dimensions as frequency or wavelength are varied and to record the neural discharge frequency as the dependent variable There is no a priori reason to assume however, that changes in the rate of response or the magnitude of response within the nervous system will bear a direct linear relation to effective magnitude of the physical stimulus The nature of the relation between stimulus magnitude and neural response is a subject of study in itself

Electrical Stimulation. In man as well as animals, direct electrical stimulation of sensory systems has sometimes been employed as a method of bypassing the primary receptor process This may provide one useful way in which limitations of the sensory responses imposed by the peripheral receptor can be revealed in man It is difficult or impossible, however, to replicate the qualitative nature of a natural or appropriate stimulus by electrical stimulation This therefore renders conclusions based on such investigations of dubious value In one area where the procedure would seem to be appropriate, the study of the temporal resolution capacity of the visual system, results with rapidly repeated electrical stimuli could be perceived as varying in time at higher rates than the maximum rate perceived as "flickering" when photic stimulation was employed In animal preparations, in addition to its value as a method of bypassing the primary receptor, electrical stimulation has also been used in an effort to determine the significance of certain efferent pathways associated with various sensory systems The role of efferent pathways is undoubtedly of great importance, but it is not clearly understood in any of the sensory systems where efferent pathways have been demonstrated anatomically

Surgical Intervention. One test of the significance of certain components in the visual system which can readily be applied with animals but never with men, at least on an experimental basis, depends upon the removal or destruction of the component. Various regions of the brain may be ablated or destroyed Nerve tracts may be sectioned Animals trained to respond to stimulation in a given way prior to such procedures may then be tested to see whether they have retained their prior ability If they have not, the conclusion that the component in question or the pathway in question is essential to the kind of discrimination under study may be justified It is always important, however, to carry out retraining procedures Sometimes surgical intervention may disrupt behavior only temporarily The demonstration of relearning ability of a given sensory discrimination may reflect such temporary disruption, or it may reveal the presence of alternative mechanisms which can be called into play following damage or destruction to the primary mechanisms which subserve the response in a normal animal Such experiments are extremely important for our understanding of sensory processes, but they are extremely difficult to interpret and must be carried out with great care and with careful attention to experimental controls

One surgical procedure that provides a measure of control within a single organism is the split-brain technique pioneered by Roger Sperry (Sperry, 1961) The optic chiasm, the corpus callosum, and other cerebral commissures are sectioned medially Communication between the right half of the brain and the left half of the brain is thereby presumably eliminated When a visual sensory discrimination is then studied by stimulation of the eye on one side this discrimination will be mediated by the visual nervous system on that same side The effects of surgical intervention at certain locations within the visual system on one side may be observed and compared with performance of the animal in the same kind of discrimination task when the other side is employed A lack of communication between the two halves of the system is supported by studies which show no transfer of discriminations learned on one side in subsequent testing of the other side The situation however, is not as clear-cut as one might hope Studies that have failed to demonstrate any transfer have been based on the use of food as a reward for correct performance Other studies have proven that when punishment is used to induce performance, there may be transfer, even when positive food rewards have failed to elicit any evidence of transfer (Sechzer, 1964) As is true especially with any form of surgical intervention, care must be exercised in the interpretation of results of experiments with the split-brain preparation

Although it is not possible to create a split-brain experimentally in man, sometimes the same effect is achieved as a result of accident or a tumorous growth which must be removed surgically Although the number of human "split-brain" subjects is limited and the precise nature of the split cannot be determined in most cases until long after experiments are performed, observations of such individuals have provided useful information concerning the laterality of the cortex in the processing of certain sensory information and the mediation of certain kinds of behavior This is particularly helpful in the studies of language (Gazzaniga, 1967)

Computer Analysis. In recent years, investigators of sensory processes have become increasingly concerned with the detailed nature of temporal patterns of neural response Whereas such patterns can be studied only with the greatest difficulty from kymograph records, they can be subjected to elaborate analyses with the aid of digital computers To date, the results of complex computer analyses have not been of far-reaching importance, but it seems likely that ultimately computers will play a highly significant role in the understanding of sensory processes in particular and the functioning of the entire central nervous system in general

NEURAL RESPONSE TO STIMULATION

Graded Potential The nature of the initial response of a receptor system to stimulation varies with the sensory modality In most instances, the initially detectible electrical response is in the form of a graded potential That is, the amplitude of the response varies, or is graded, dependent upon the amplitude of stimulation Such a response may be relatively sustained during the time course of stimulation, although frequently it will show an initial maximum level which may reduce somewhat with continuation of a sustained stimulus The origin of such graded potentials is not clearly understood, although they may arise in primary receptors They differ from the spike discharge of neurons in that their amplitude may vary, and that they are sustained in duration rather than occurring repetitively with sustained stimulation Polarity of the graded potential may vary dependent upon the nature of the stimulus (Fig 8 1). Records obtained in the region of the retinal receptors show a difference in the polarity of the response dependent upon the wavelength of the stimulus (Mac Nichol and Svaetichin, 1958) Only in the case of the olfactory system does the primary receptor serve also as the primary neuron, discharging with impulses, the frequency of which varies with the concentration of the stimulating substance

In the visual system, as detailed in a subsequent chapter, there is evidence that initial responses at the receptor and cells associated directly with the receptor may be a hyperpolarization Hyperpolarization in one cell apparently can give rise to depolarization of a cell with which it makes contact and subsequent neural discharge of that cell which is propagated along an axon

The Generator Potential. Before a cell will fire there must be a depolarization of the cell membrane to a certain "threshold" level When this level is reached, the spike discharge follows In the wake of a spike discharge, the sustained stimulation of the cell by contacting elements results in a repetition of the process If the depolarization does not reach the re-

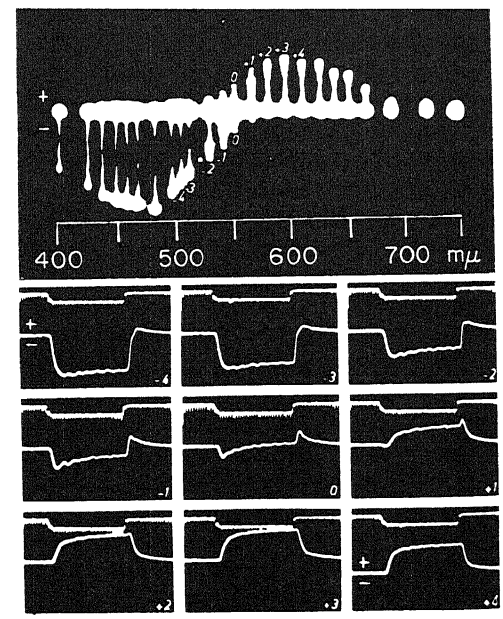


Fig 8 1 Microelectrode records of S (slow) - potentials in the retina of the fish The upper portion of the figure indicates the relative polarity of response as a function of the wavelength of the stimulus Lower records show the time course amplitude and polarity of responses to 300 msec stimulation The stimulus is marked by a slight downward deflection of the timing trace (MacNichol and Svaetichin 1958)

quired level, the cell will not discharge (Fig 8 2) It is evident that as level of stimulation is increased, the generator potential reaches the critical amplitude more rapidly, permitting a more rapid frequency of discharge of the cell The spike discharge occurs on top of the measurable generator potential The generator potential is monophasic, of relatively short latency and, as suggested above, may be graded in amplitude It thus is distinguished from the propagated action potential which is of roughly constant amplitude for a given nerve fiber

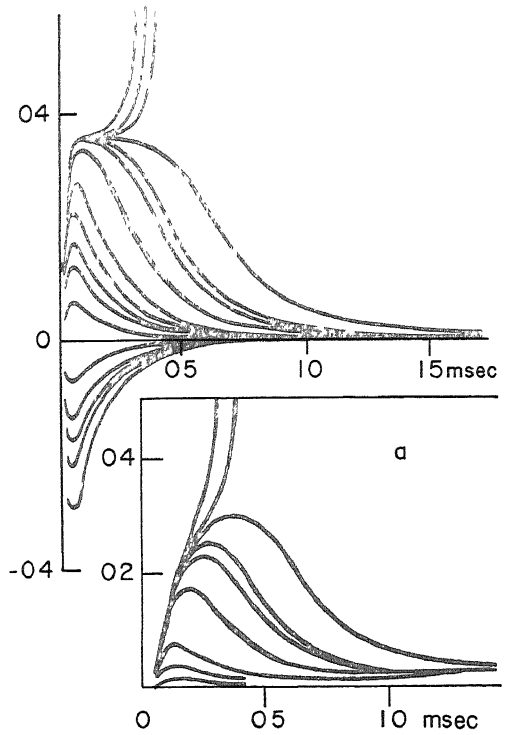


Fig 8 2 Generator potential resulting from cathodal shock to large diameter crab nerve fiber Successively higher amplitude curves result from increase in shock strength At a shock strength of slightly less than 15 mv the amount of membrane depolarization requisite to spike discharge is reached and the cell fires (After Hodgkin 1938)

Nerve Fiber Discharge. The pattern of a nerve fiber's discharge frequency may vary considerably depending upon the nature of stimulation as well as the specific modality under consideration With the initiation of a stimulus, spike discharge frequency may be very rapid after a short latency, followed by a sustained discharge frequency at a higher rate than the resting level until the stimulus is terminated Upon termination of the stimulus, there may be a brief period of inhibition of any response prior to resumption of some resting level of response Some fibers may be quiescent and show no detectible resting level of discharge The relation of discharge frequency to intensity of stimulation is illustrated in Figure 8 3 for a single optic nerve fiber of the horseshoe crab It is evident that as the stimulus intensity is reduced, latency of response grows longer, the initial frequency of discharge is lower, and the repetitive discharge of the unit, while the stimulus remains on, becomes slower and less regular

By reason of interaction processes within sensory systems, not all cells respond with heightened discharge activity upon stimulation Some may be inhibited by a stimulus (Fig 8 4) The top trace shows a unit which

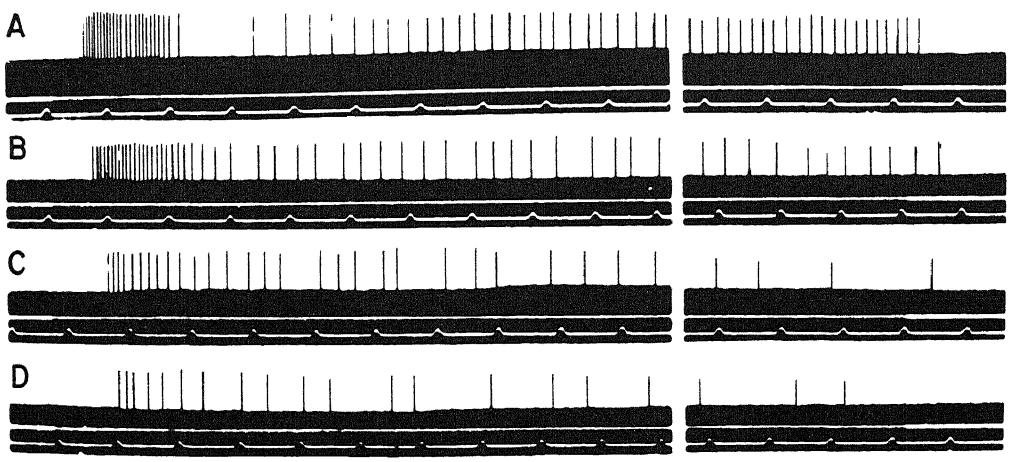


Fig 8 3 Discharge of a single optic nerve fiber of *Limulus* for stimulation at each of four levels of stimulus illumination A 6 3 × 10^4 B 6 3 × 10^3 C 6 3 × 10^2 and D 6 3 × 10 meter candles (Hartline and Graham, 1932)

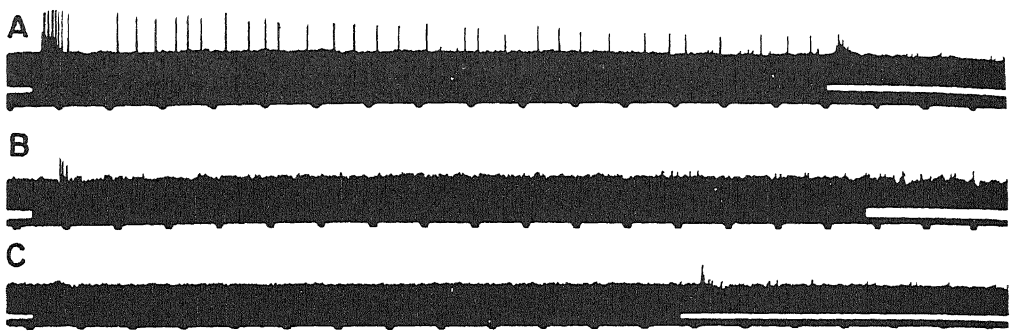


Fig 8 4 Three different types of discharge from retinal fibers of the frog Stimulus duration is shown by the darkened portion of the white line above the fifth-second time marker (Hartline 1938)

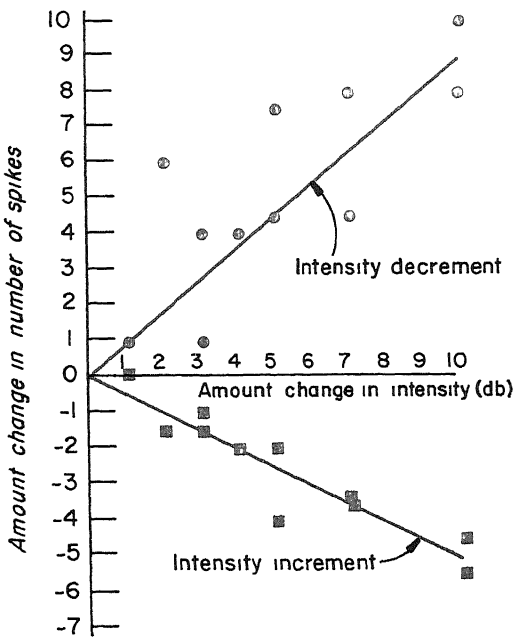


Fig 8 5 Changes in number of spike discharges associated with brief changes in stimulus intensity for a monkey lateral geniculate cell (DaValois Jacobs and Jones 1962)

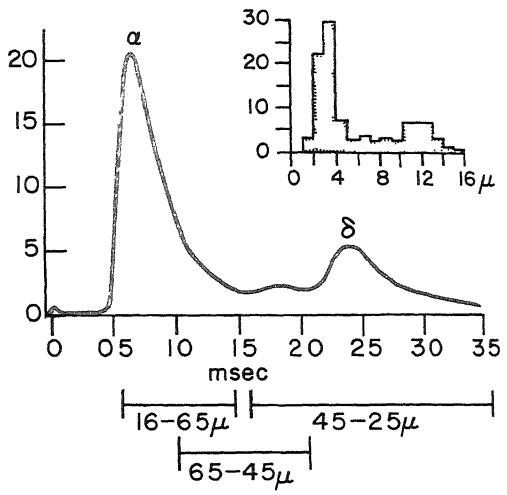


Fig 8 6 Compound potential in a human sensory nerve voltage (arbitrary units) as a function of time Ranges of fiber diameters are shown below the time axis (Gasser 1960)

responds initially with a rapid burst of activity and then a sustained level of activity while the stimulus remains on The middle trace illustrates a unit which responds with a burst of activity upon initiation of the stimulus, relative quiescence while the stimulus is maintained, and absence of activity upon termination of the stimulus The bottom trace shows a unit which responds not at all but rather is inhibited when a stimulus is presented Upon termination of the stimulus, a burst of activity occurs and is sustained for some time prior to return of the unit to its resting level

In the visual system, certain ganglion cells may respond with increased activity upon illumination of a certain region of the retina and decreased activity as illumination is reduced There are also cells which respond directly opposite to this Activity increases with a decrement in intensity of the stimulus and decreases with an increment in the stimulus The activity of this kind of a unit as a function of the amount of change of the level of stimulus intensity in decibels (ratio steps of 1 26) is illustrated in Figure 8 5 Obviously, for any cell to show a decrement of response to some form of stimulation, as compared with the absence of any stimulation, there must be a resting level of activity which can be decreased

Neural Coding of Stimulus Information. It is evident that information is encoded in neural activity in terms of the rate of that activity Changes in stimulus intensity are reflected by increase or decrease in spike discharge frequency Complex patterns of temporal change of stimulation are probably also reflected in terms of the temporal pattern of spike discharge The possibility of a temporal encoding in the spike discharge of other than temporal aspects of the stimulus dimension has been considered but cannot be supported by substantial experimental evidence The most important aspect of coding of information transmitted by neurons seems to be that of the identity of the neuron The nature of the messages in terms of their spike discharge frequency is remarkably similar from one sensory modality to another Apparently, the specific neurons which carry the message, by reason of their identity, their origin and their termination, relate to specific qualitative aspects of the stimulus as these are interpreted by the brain

Compound Action Potential. When a bundle of nerve fibers is transmitting information in response to sensory stimulation, a compound action potential may be recorded from the bundle with a gross electrode This shows changes in voltage with time which reflect the summed electrical activity of all of the components within the bundle When a large number of fibers is firing simultaneously or at nearly the same time, the amplitude of the action potential will be high Compound action potentials typically show several peaks, associated with classes of neurons having roughly comparable conduction velocities Thus, a wave of activity for a given class of fibers reaches the electrode at very nearly the same time for all of the fibers within the class These fibers are then represented by a specific component of the compound response (Fig 8 6)

The Physics of Sound and the Biophysics of Hearing

Hearing rivals vision as a means of deriving highly complex and detailed information from the environment In humans, endowed with speech, it is the most important channel for interpersonal communication Hearing is the detection of vibrations transmitted through an elastic medium It has evolved from vibration-sensing organs, examples of which may be found in many species more primitive than man and the higher animals The term "sound" is usually reserved for those frequencies of vibration to which our hearing apparatus will respond, but if we generalize even over the mammalian species, this covers a range of frequencies much broader than those that can be detected by man Differences lie in the structure of the detecting apparatus, its size and physical characteristics In the higher animals, the differences in most instances may be dismissed as quantitative rather than qualitative, and much of our knowledge of the physiology of hearing is derived from infrahuman species such as the cat and the guinea pig

THE PHYSICS OF SOUND

The physical stimulus for hearing is a longitudinal pressure wave consisting of periodic compression and rarefaction, which requires an elastic medium for its transmission The frequency of sound is determined by the originating source and remains invariant as the sound is transmitted through various media Velocity and wavelength are related by the expression $V = f \times \lambda$, where V represents velocity, f is frequency, and λ stands for wavelength

The velocity of sound in air is approximately 330 meters/sec, while in water it is nearly 1500 meters/sec The way in which the propagation of sound is impeded in a given material varies with frequency Impedance refers to the attentuation of amplitudes and not to velocity One component of impedance increases with the elasticity of the material, but varies inversely with the frequency of the sound On the other hand, the higher the mass of a given material, the higher its contribution to impedance This contribution increases directly with frequency These two reactive components of impedance, that contributed by the elasticity (s) and that by the inertial mass (m)—characteristics of the material—are of opposite sign and add algebraically The third component of impedance is the resistive component, this is added vectorially to the algebraic sum of the reactive components It is independent of frequency The complete expression for mechanical impedance, Z, is

$$Z = \sqrt{[2\pi fm - (s/2\pi f)]^2 + R_m^2} \qquad (1)$$

where R_m is the frictional resistance in ohms The total impedance is a minimum when the reactive component $2\pi fm - (s/2\pi f)$ is a minimum This will be the case at that frequency where these components are equal in absolute value and hence will cancel each other out Thus, by setting $2\pi fm$ equal to $s/2\pi f$ one can solve for frequency where the reactive component is 0 The resulting equation is

$$f = \frac{\sqrt{s/m}}{2\pi} \qquad (2)$$

This provides a method of calculating the natural frequency of a given material, that frequency for which the attenuation of the material is minimal

Sound may be transmitted from one medium to another, but the efficiency with which this occurs will vary depending upon physical characteristics of the media The more nearly their impedances are the same, the greater will be the efficiency of transmission It is important to have some knowledge of the effects of impedance and how this changes with materials and with frequency to understand the auditory mechanism

Sound is vibration, and vibration may be described in terms of its frequency and ampli-

tude The most interesting and important sounds, however, are not simple vibrations Speech and music, for example, are exceedingly complex, and the information conveyed by such sounds is in no way uniquely dependent upon specific frequencies or even frequency complexes Rather, it is the nature of relations among the frequency components of the complex sound and the way these are distributed in time which render music recognizable and enjoyable, or speech intelligible

The fact that complex sounds can be considered to consist of sinusoidal components provides some justification for the use of simple sinusoidal stimuli in the study of the auditory mechanism The ability of the ear to discriminate one sinusoidal frequency from another may be of relatively small importance for normal communication, but it is this basic capability which permits the recognition of complex sounds from which important information may be extracted Single sinusoidal frequencies of pure tones and brief clicks have been the most frequently used stimuli in the study of audition A brief click may be characterized as a broad band of frequencies compressed into a narrow interval of time "White" noise has also been used It is so named by reason of the fact that it consists, as does white light, of a broad band of frequencies which are of approximately uniform amplitude as well as of random phase relation

THE DIMENSIONS AND SPECIFICATION OF SOUND

The *frequency* range of sound to which the human ear responds is from 16 or 20 Hz (Hertz, or cycles/second) to 20,000 Hz There is a gradual reduction in the upper limit of this range with increasing age The range of sound *intensities* to which the ear can respond normally is extremely broad At the frequency of maximum sensitivity, between 1000 and 3000 Hz, the ratio of maximum intensity that can be heard without damage to the ear, to the minimum intensity that can be detected, is approximately 10^{12} The minimum power to which the ear can respond is of the order of 10^{-16} watt/cm ^{2} This is a standard reference value, in fact, many individuals can detect lesser amounts of power This power level is comparable to a sound pressure of approximately 0 0002 dyne/cm ^{2} at 20° C

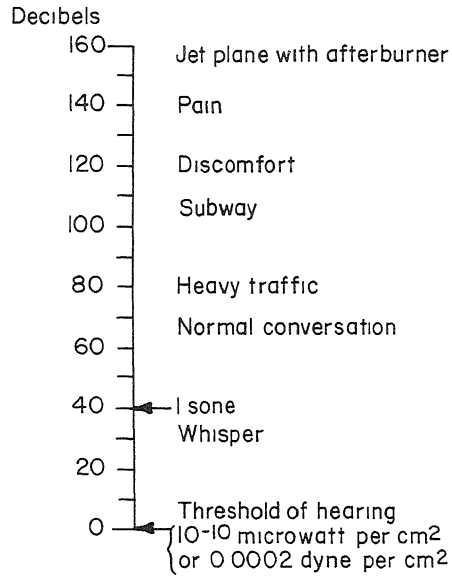
The tremendous range of physical energy to which the ear responds as well as the nature of the response of the auditory system renders the use of a logarithmic scale desirable for the specification of sound intensity Such a scale represents increases in ratio steps The unit used in audition is the decibel A unit of 1 bel (named after Alexander Graham Bell) represents a ratio increase by a factor of 10, 2 bels a factor of 100, etc There are thus only 12 bels over the range from approximate threshold to the level at which sound becomes painful in the frequency range of maximum sensitivity The decibel (db) was devised as a unit which would provide a finer index of sound level As the name implies, there are 10 db to the bel, but this obviously does not mean that a decibel represents a ratio increase by a factor of one The meaning of the decibel is best understood by illustration of the method of computation

$$\text{No of db} = 10 \log (J/J_0) \qquad (3)$$

The logarithm to the base 10 of the ratio of a given power level, J, to some reference level, J_0, provides a statement as to the number of bels J is above or below the reference level By multiplying this logarithm times 10, the answer is converted to decibels The number 10, by which the logarithm is multiplied, is also the logarithm to the base 1 26 of 10, that is, 1 26 raised to the 10th power equals 10 By multiplying a logarithm to any given base (e g , 10) by the logarithm of that base to some other number as a base, the logarithm is transformed to that for the other number as base Decibels are thus logarithms to the base 1 26 and 1 db represents a ratio of 1 26

Sound intensity is properly a specification in terms of *power* level The power J is proportional to the sound pressure times the velocity of a particle in the medium through which the sound is being propagated The latter, in turn, is directly proportional to pressure Thus, sound power is proportional to pressure squared If one sound has a pressure 10 times as great as another, they are said to differ by 20 db The ratio of the pressures is only 10, but the ratio of the powers of the two sounds is 100 and the difference between them in decibels is therefore 20

The decibel notation has been applied in comparing electric powers, voltages, and currents. It is a useful notation, but it must be remembered that it is an arbitrary one and



Graph 1 A decibel scale showing approximate intensity levels produced by various sources The sone is a unit of loudness based on ratio estimates The reference for such estimates is the loudness of a 1000-Hz tone at 40 db (Stevens 1958)

that its relation to actual physical values depends upon a knowledge of the reference level employed Standard reference levels used frequently include 0 0002 dynes/cm ^{2}, or 10^{-16} watts/cm ^{2} The former is the standard reference adopted by the American Standards Association, as mentioned above, the two levels are equivalent at 20° C They represent a level of sound near threshold for the frequency of maximum sensitivity in the normal ear A reference level of 1 dyne/cm ^{2} is also used by some investigators In addition, decibel notation may be employed to represent the amount of attenuation of a signal below the maximum level that can be obtained from a given amplifier This is appropriate in order to afford a relative scale in many experimental procedures The numerical values are quite arbitrary, however, so far as their relation to any fixed physical standard is concerned, and this must be recognized in interpreting data expressed in this fashion The levels in decibels of some typical sources of sound are shown in Graph 1 relative to the standard reference The *sone* is a unit of loudness based on a subjective ratio scaling technique

ANATOMY AND PHYSIOLOGY OF THE EAR

The human ear functions in air, and, understandably, represents a fairly efficient device for the reception of sound transmitted through air When functioning normally, it is stimulated by pressure waves transmitted through air and along the external auditory canal, or external meatus, to the eardrum, or tympanic membrane Here, the medium of transmission changes but sound continues to be transmitted in the form of mechanical vibratory energy up to the point where the mechanisms of the inner ear are stimulated At that point there is a conversion of the stimulating energy to neural activity in the receptors which permits the discrimination of sounds that differ in character As mentioned earlier, the ear responds to frequencies between approximately 20 and 20,000 Hz It is not equally responsive to all of these frequencies, however, as illustrated in Figure 8 7 Optimum response occurs at approximately 2000 Hz The nature of this frequency response is determined in large measure by the physical characteristics of the receiving organ, the nature of the external auditory canal, the tympanic membrane, the chain of small bones which link the tympanic membrane through the middle ear to the fluids of the inner ear, and finally the coupling of sound into these fluids which surround the receiving mechanism, the basilar membrane

The external ear, or pinna, serves some function in channeling energy into the ear and in influencing the directional response of the ear It is not nearly so important in man as it is in lower animals In some animals with highly

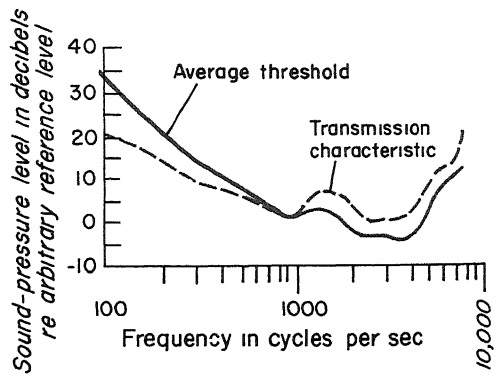


Fig 8 7 Transmission characteristic of the ear based on the summed impedances of the tympanic membrane malleus and incus the middle ear cavities and the stapes cochlea and round window membrane as calculated by Zwislocki Comparison with an average threshold curve illustrates the important influence of the physical characteristics of the auditory mechanism on the frequency response function in hearing (Zwislocki 1965)

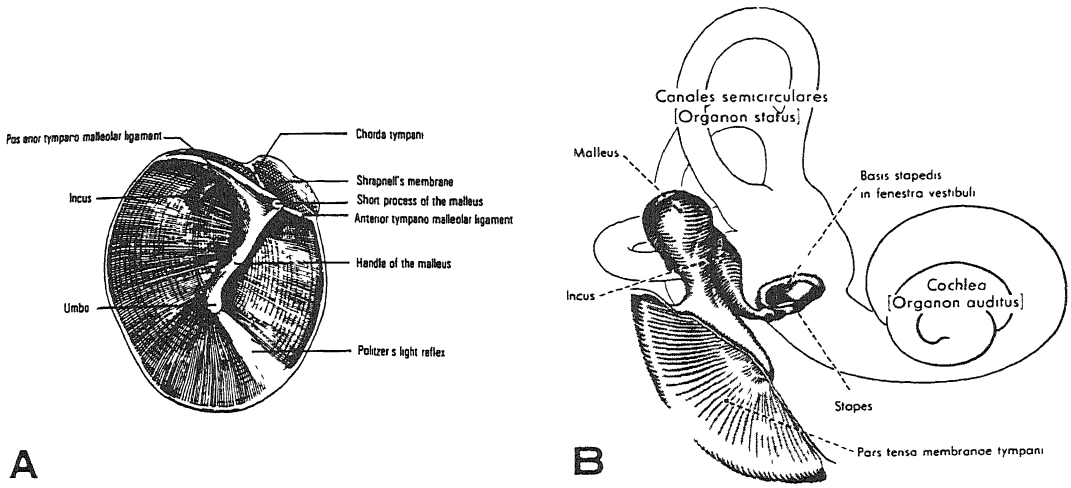


Fig 8 8 *A* Lateral surface of the right tympanic membrane as seen with the otoscope (Portmann 1951) *B* Cut away section of the tympanic membrane showing the nature of attachment of the malleus along with its linkage via the incus and stapes to the oval window of the labyrinth (Mellonı 1957)

refined motor control of the external ear, it is an important factor in the localization of the origin of sound Sound is transmitted along the external meatus to the tympanic membrane which is set in motion by the sound vibrations Its total area is approximately 70 mm ^{2} The first of the bones of the middle ear, the malleus, is attached to the inside of the tympanic membrane by its footplate The appearance of the membrane and the point of attachment of the malleus may be visualized in Figure 8 8 The malleus is connected via an intermediary linking bone, the incus, to the stapes, which in turn is attached to the oval window of the cochlear mechanism The bones of the middle ear are surrounded by air which is ordinarily at approximately the same pressure as the external atmosphere Pressure in the middle ear is equalized via the eustachian tube which extends from the middle ear to the pharynx If one calculates the lengths of the bones in the ossicular chain and considers the way in which they are articulated, the mechanical advantage from tympanic membrane to oval window appears to be approximately 1 3 This is probably of relatively little importance, however, and may serve merely to offset the frictional resistance which exists along this pathway

Up to the level of the tympanic membrane the receiving organ is a completely passive organ At this point, however, it may be influenced internally The articulation of the small bones of the middle ear can be influenced by the middle ear muscles, the tensor tympani and the stapedius Increased tension in these muscles reduces the efficiency of transmission of low sound frequency (<1000 Hz) along this pathway Contraction may therefore serve as a protective device, and in fact, it can be demonstrated that contraction of the stapedius muscle occurs as a reflex response to loud sounds As in the pupillary reflex of the eye, this response is consensual so that contraction in both ears is elicited by stimulation of only one ear This system thus provides a feedback loop for control of the auditory receptor It may afford some protection against loud sounds, but the latency of the reflex is too great for this to be of major significance A more important function may be in the reduction of the masking effects of low frequency sounds Low frequencies are particularly effective in the masking of other frequencies, but relatively unimportant with respect to information content

The footplate of the stapes connected to the oval window is approximately 3 mm ^{2} in area When this is considered in relation to the *effective* area of the tympanic membrane, approximately 40 mm ^{2}, it is evident that there is a mechanical advantage of 13 to 1 between the tympanic membrane and the oval window This provides for a reasonable impedance match between the external medium of propagation, air, and the internal medium of propagation, the perilymph of the cochlea

Sound waves are transmitted very rapidly

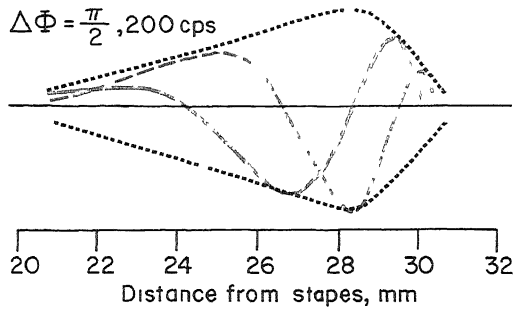


Fig 8 9 Traveling wave along the cochlear partition for a 200 Hz tone The solid line indicates the deformation pattern at a given instant The line with the short dashes shows the same traveling wave one-fourth of a period later The envelope shows the maximum displacement at each point (von Bekesy 1947)

through the perilymph of the vestibular canal around the approximately 2¾ turns of the cochlea to its tip, then through the helicotrema, the connecting point between this canal and the tympanic canal which carries the sound vibrations down to the round window As the oval window bulges inward, the round window bulges outward in compensation The basilar membrane which lies between these two canals is thus surrounded by liquid through which a sound wave is being propagated There is very little delay between the arrival of sound pressure increases at the oval window and the bulging out of the round window in compensation Nonetheless, the actual transmission of vibrations along the basilar membrane caused by the stimulating sound can be shown to be somewhat delayed with respect to the transmission of sound through the liquid medium of the inner ear The reason for this is that sound is coupled into the basilar membrane almost solely at its base The physical characteristics of the basilar membrane are such that it is relatively stiff at the base and becomes less so along its length toward the apex The change in stiffness is by a factor of 100 along the entire length At the same time, the membrane becomes broader toward its apex This is true even though the cochlea as a whole is becoming narrower Thus, the increased breadth of the basilar membrane occurs at the expense of the size of the liquid-filled canals

Observations of the basilar membrane have been made with stroboscopic illumination in order to determine the nature of its motion in response to sound stimulation It is fairly well established that the membrane is set in motion at the base and that traveling waves move out along the membrane from the base toward the tip The reason for this is the efficiency of the coupling of sound by reason of impedance matching between the perilymph and the base of the basilar membrane and the relatively broad frequency response characteristics of the base Vibrations are transmitted out along the membrane but are attenuated as they approach its tip The amount of attenuation increases with the frequency of the sound As the membrane becomes less stiff, higher frequencies of sound are damped to a greater extent

There are not fixed, individual points along the membrane which, at a given frequency of stimulation, represent maximum amplitudes of excursion or minimum amounts of excursion, i e , nodal points, rather the point of maximum amplitude progresses along the membrane out to a distance which is determined by the frequency of the stimulus The lower the frequency of the driving sound pressure wave, the greater the proportion of the membrane which is set in motion For frequencies of up to about 20 to 30 Hz the entire membrane is set in motion and it vibrates in phase That is, all along its length it is moving either up toward the vestibular canal or down toward the tympanic canal at the same time For frequencies of up to about 200 Hz , the entire membrane is set in motion but there will be at least one phase change along its length Part of it will be moving in one direction and another part in the opposite direction As suggested earlier, the wave will move along the membrane as the motion occurs For frequencies above the 200 Hz level, less than the entire membrane is set in motion and the higher the frequency, the less will be the length of the membrane which moves The envelope of the traveling wave which moves along the membrane is relatively constant for a given condition of stimulation That is to say, the maximum aplitude which any given point on the membrane will reach is a constant, although that point may sometimes be moving with its maximum amplitude and at other times may barely move at all These ideas are illustrated in Figure 8 9 The envelope of the traveling wave is shown and the nature of the excursion of the basilar membrane at any given point is illustrated by the lines drawn within the dashed lines which rep-

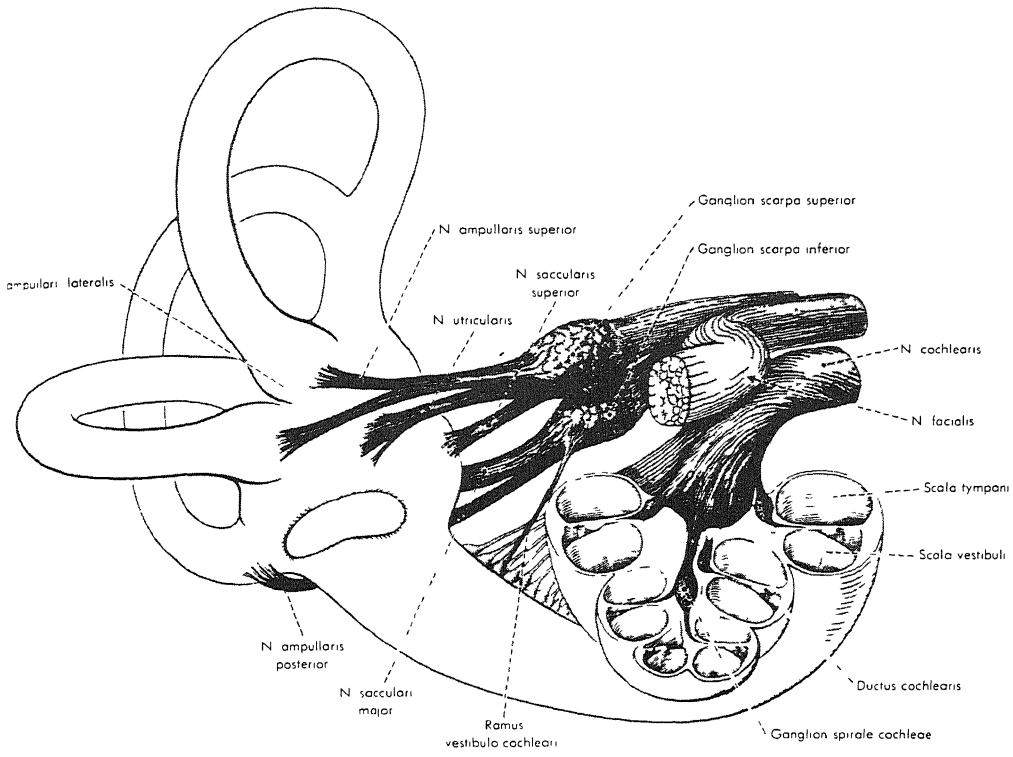


Fig 8 10 A midmodiolar section through the right cochlea shown in relation to the vestibular system with vestibular and cochlear divisions of the eighth nerve (Melloni 1957)

resent the limiting envelope The short-dashed line and the solid line represent two different points in time and two different degrees of progression of the traveling wave along the membrane It is thus clear that the nature of the motion pattern of the basilar membrane is determined by the frequency characteristics of the sound stimulation The motion characteristics are not simple, however, and even though one point along the membrane may represent a greater maximum amplitude of excursion than any other for a given single frequency of stimulation, the way in which it is moving will vary with time during the course of its stimulation

In summary, virtually all of the energy transmitted to the inner ear is absorbed by the basilar membrane at its stiff basal end, and motions of the basilar membrane are propagated along its length toward the broader, less stiff apex most efficiently for lower frequencies that are less damped by this organ The nature of the motion pattern is a characteristic of the stimulating frequency, hence, the pattern of excitation of receptors distributed along the length of the membrane is also determined by the nature of the stimulus The nature of the motion will be complex for a complex stimulus, but will be unique and characteristic of any unique stimulus complex For a given frequency of stimulation, the nature of the motion pattern will be altered by the amplitude of stimulation The greater the amplitude, the greater will be the amplitude of excursion of the basilar membrane and the greater will be the length of the membrane involved in the motion for those frequencies higher than that for which the entire length of the membrane is all ordinarily set in motion As illustrated in Figure 8 9, the envelope of the traveling wave shows how damping results in the falling off of amplitude of motion to zero beyond a certain region of the membrane at a given frequency The nature of the motion of the basilar membrane follows from its physical characteristics Actually, observations by von Békésy (1960) of the motion provided important clues to the physical characteristics of this organ

How the system permits discrimination of one frequency from another has been the sub-

ject of extensive debate, but it is clear that the pattern of motion of the basilar membrane is characteristic for a given sound It is reasonable to assume that the ability of the system to analyze sound is a result of both temporal and spatial patterns of excitation of the receptors along the length of the membrane For a certain frequency, a certain part of the membrane is driven to motion of the greatest amplitude, but it might be an oversimplification to suggest that those receptors in the region of maximum amplitude are exclusively responsible for the detection of sound of that frequency This problem is discussed further below

TRANSFORMATION OF ENERGY

One of the most interesting problems in understanding the auditory system is that of how mechanical energy of motion of the basilar membrane is converted into neural activity To pursue this question it is necessary to examine the anatomy of the cochlear system in some detail The cochlea is shown in Figure 8 10 in its relation to the rest of the labyrinthine mechanism The cross section of part of a single turn of the cochlea is illustrated in Figure 8 11 The tectorial membrane is connected to tissue which is in turn attached to

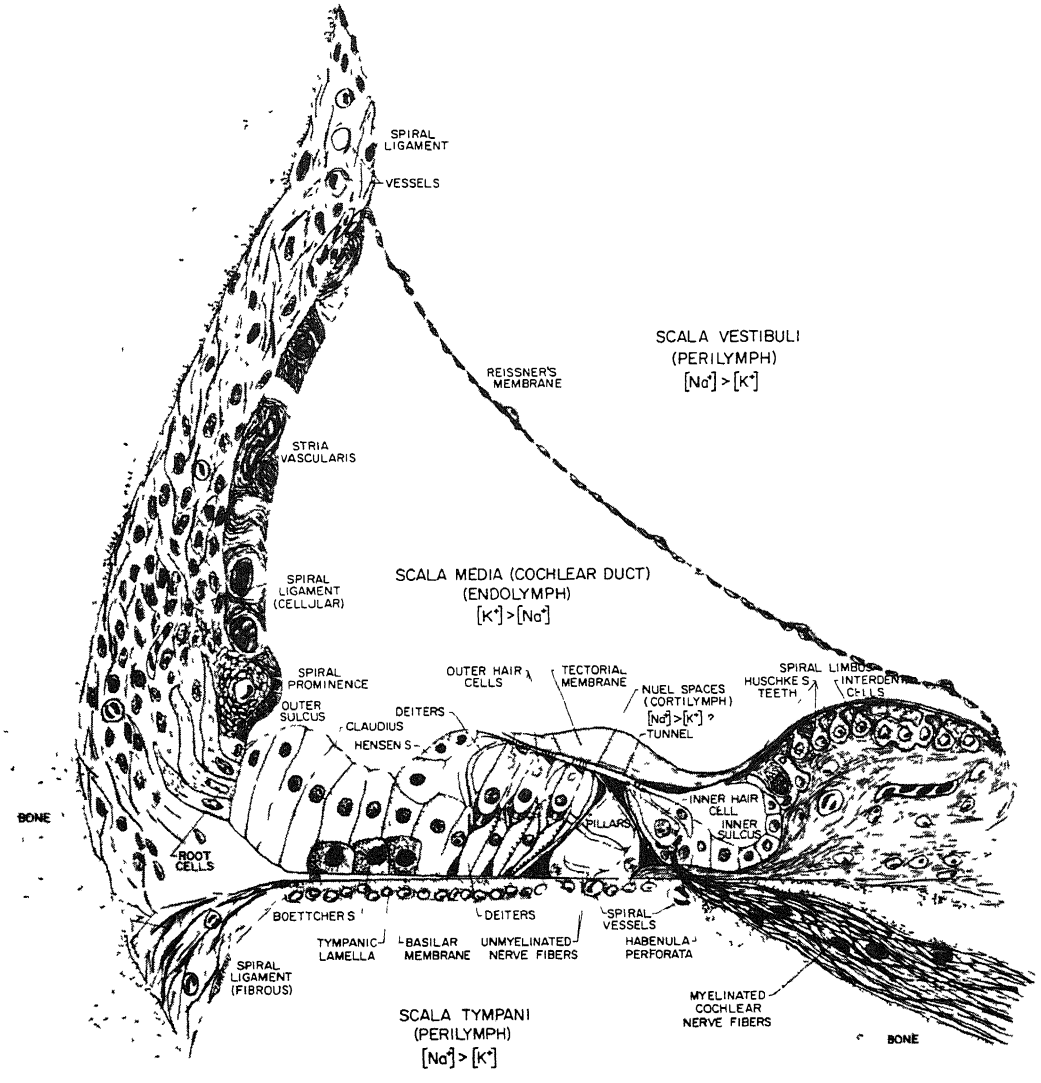


Fig 8 11 The major cochlear structures as seen in a transverse section of the cochlear duct from the upper portion of the first coil of the guinea pig

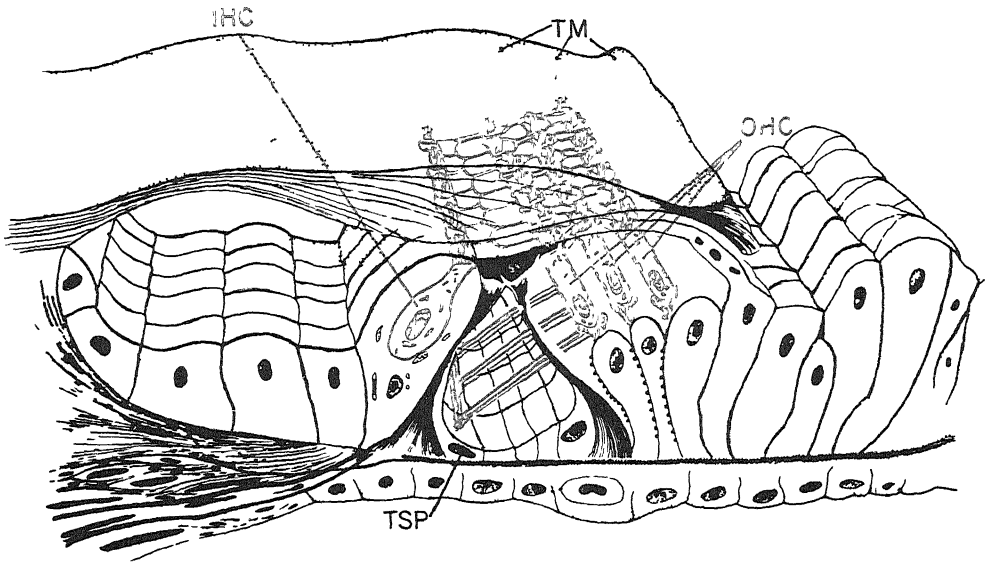


Fig 8 12 Schematic illustration of the structure of the mammalian organ of Corti OHC outer hair cells IHC inner hair cells TM tectorial membrane TSP tunnel spiral fibers (Wersäll Flock and Lundquist 1965)

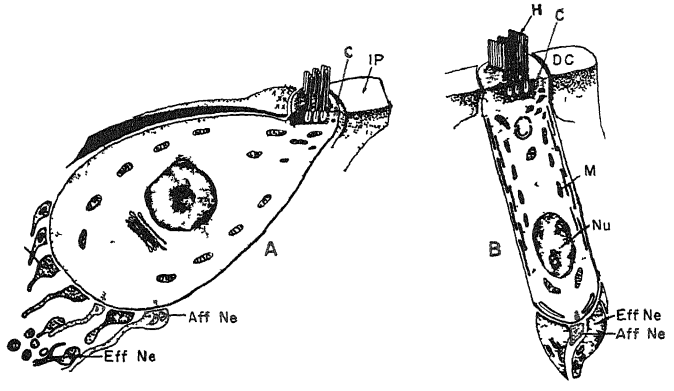


Fig 8 13 Relative shapes of the inner (*A*) hair cell and the outer (*B*) hair cell C centriole or basal body Aff NE afferent nerve ending Eff NE efferent nerve ending IP inner pillar DC Deiters cell M mitochondrion Nu nucleus (Wersall Flock and Lundquist 1965)

the central spiral laminar bone This membrane extends out over the hair cells which in turn arise from the cells of Deiters on the basilar membrane

The tectorial membrane is not attached to the hair cells and may move relative to them Such relative motion causes bending of the hair cells and it is the bending which initiates neural activity The hair cells are the primary sensory elements As illustrated in Figures 8 12 and 8 13, outer and inner rows of hair cells are separated by pillar cells or the rods of Corti which form a tunnel or arch The height of this formation increases from the base to the apex The single row of inner hair cells is found on the side of the cochlear spiral toward the center, the three or more rows of outer hair cells are on the outer side of the arch Just inside of the inner hair cells is located the inner sulcus and medial to this is a spiral ridge, the limbus spiralis, to which the fibers of the basilar membrane are anchored Reissner's membrane (Fig 8 11) is attached to the inner margin of the ridge and the tectorial membrane, which overlaps the hair cells, is attached to its surface The tectorial membrane is in contact with the hair cells but they do not penetrate its surface The hair cells are bathed in fluid but it cannot be the endolymph of the medial canal because the potassium content of

the endolymph is so high that it would be impossible for the unmyelinated fibers which innervate the hair cells to fire in such a medium The cells of Hensen lateral to the outer row of hair cells may serve as a block to prevent the mixing of the fluid around the hair cells with endolymph The nature of this fluid must be more like that of perilymph than endolymph, but its analysis is presently unknown The various constituents of perilymph and endolymph and their approximate concentrations are given in Table 8 1

Table 8 1
Composition of fluids of inner ear

	Spinal Fluid	Perilymph	Endolymph
	mEq /liter		
Potassium	4 2	4 8	144 4
Sodium	152 0	150 3	15 8
Chloride	122 4	121 5	107 1
	mg /100 ml		
Protein	21 0	50 0	15 0

After Smith, Lowry, and Wu

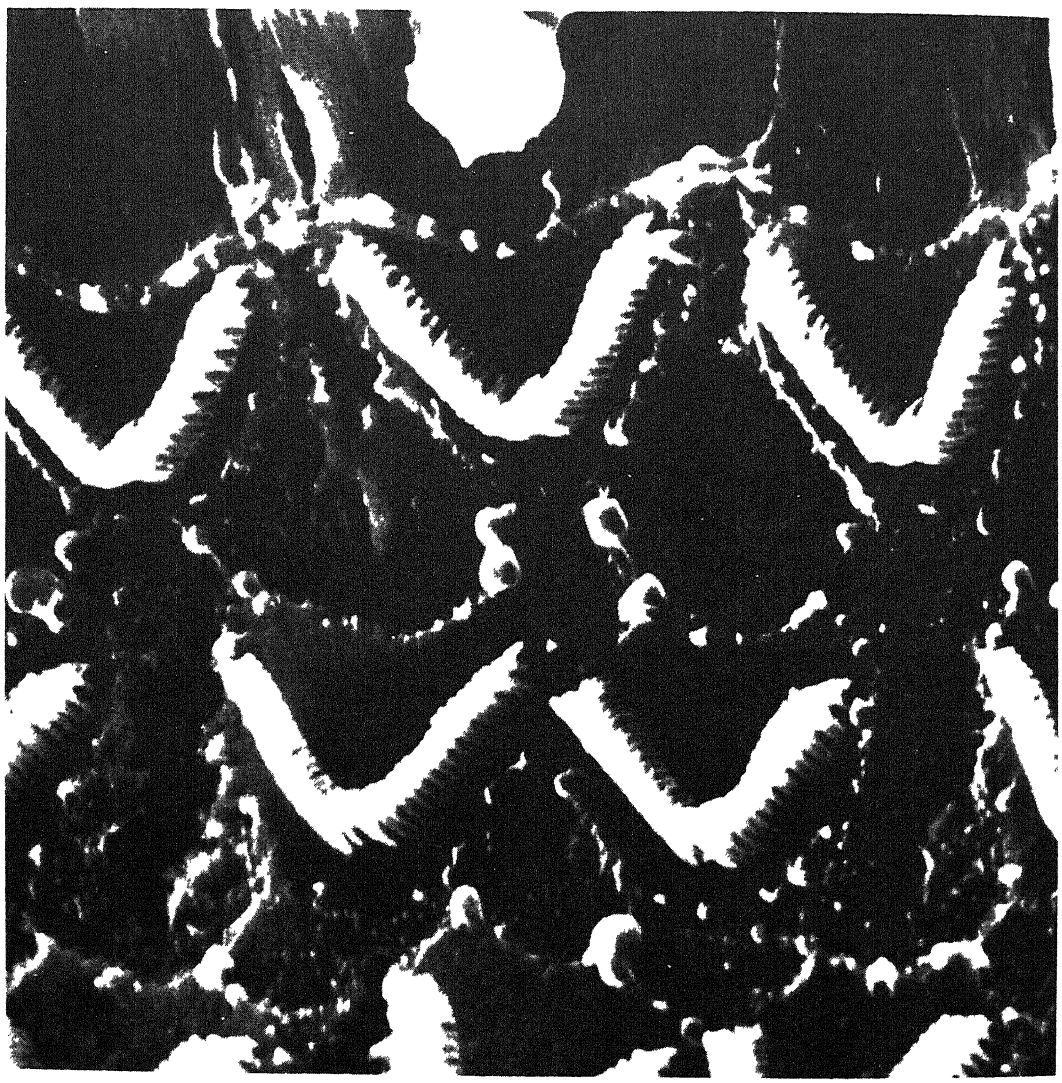


Fig 8 14 Scanning electronmicrograph of outer hair cells in man showing characteristic formation of the hairs (Photograph provided through the courtesy of Hans Engström)

It has been estimated that in man there are from 12,000 to 20,000 outer hair cells and 3500 inner hair cells in each ear (Retzius 1884, Guild, 1932, Chow, 1951) Their relative shapes are illustrated in Figure 8 13 There are as many as 120 to 140 hairs (stereocilia) on the ends of each outer cell, arranged in several rows in a W-shaped pattern (Fig 8 14) At the base of the W is a single basal body or centriole The inner hair cells differ significantly in shape, have fewer hairs (50 to 60) and also have a single basal body or centriole (Ades, Engstrom, and Hawkins, 1962) The interior structures of these cells are shown in Figure 8 13

The way in which the energy of motion of the basilar membrane is transmitted to the hair cells and they in turn transform it into neural activity is not entirely clear The basilar membrane is attached on both sides of the cochlear canal The tectorial membrane is attached only on the side adjacent to the modiolus, the center of the spiral Motion of the membrane apparently causes a rocking motion of the rods of Corti and the hair cells, and this in turn results in a shearing motion between the ends of the hair cells and the tectorial membrane Thus, as the basilar membrane moves up and down, the tectorial membrane slips relative to the tops of the hair cells and tends to bend them The precise nature of this shearing motion undoubtedly differs in different regions of the system, depending on the nature of the stimulus and its relative intensity Out near the end of the envelope of the traveling wave, the hair cells are apparently bent in a longitudinal direction along the long axis of the basilar membrane Cells located closer to the base may tend to be pressed down while other hair cells still closer to the base appear to be bent laterally (von Békésy, 1943) The probable extent of the lateral shear and the longitudinal shear relative to the envelope of a traveling wave based on calculations by Tonndorf (1960) are illustrated in Figure 8 15

The hair cells of the inner ear are analogous to other cells found in the utricular mechanism and in the lateral line organs of some species These other cells are more readily accessible and hence more easily studied They possess a kinocilium, a kind of stereobody with a projecting cilium It has been demonstrated that bending of the individual hairs, or cilia, in the

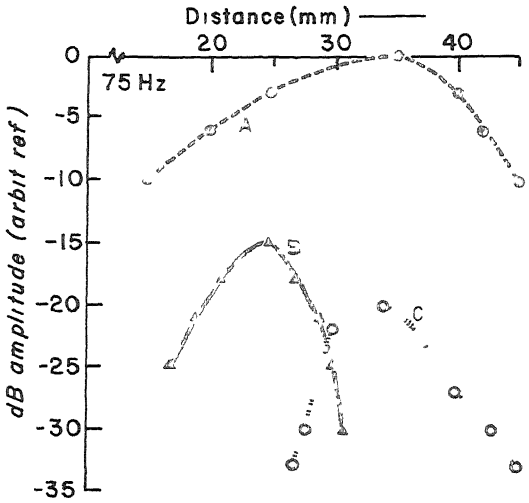


Fig 8 15 Envelopes of the three modes of motion in a cochlear model (*A*) traveling wave (*B*) radial shear (*C*) longitudinal shear The maximum amplitude of the traveling wave pattern is arbitrarily set at unity The ordinate is given in decibels of attenuation Note that the transformation of the traveling wave motion into shear motion is accompanied by an amplitude reduction by a factor of 15 to 20 db (Tonndorf 1960)

direction of the kinocilium in these other cells results in cell depolarization (Lowenstein & Wersäll, 1959) Bending in the opposite direction results in hyperpolarization By analogy, one might assume that bending of the hairs of the cells of the inner ear laterally and away from the cochlear center would cause depolarization, for it is in this direction that the basal bodies or centrioles are located The problem has not been resolved, however Von Békésy (1953) has studied the matter and reports that the inner hair cells appear to be bent longitudinally while the outer hair cells are bent laterally As mentioned above, the basal body is oriented laterally and away from the center for both outer and inner hair cells. Engstrom et al (1962) have argued that the structure of the inner hair cells is such that they should be expected to respond to lateral shear while outer hair cells might be expected to respond to bending in a variety of directions Thus, the precise nature of the mechanism of action which gives rise to the excitation of these cells is not known The hairs, themselves, must transmit the mechanical motion to the interiors of the cells The problem in interpretation of how the hair cells work rests in part with their incredible sensitivity At threshold,

based on measurements of movement of the malleus (Wilska, 1935) and stapes (von Békésy, 1960), movements of the basilar membrane must not be very much greater than 10^{-9} centimeters This is less than the thickness of the plasma membrane of the hair cells themselves

It is difficult to discern what kind of mechanism permits the conversion of such minute mechanical displacement into a significant amount of neural activity One point is clear, the process is an active one The energy of the activity elicited by stimulation greatly exceeds the energy of the stimulus at threshold The receptor itself, dependent upon its active metabolic processes, contributes significantly to the response This has been demonstrated elegantly by experiments performed by von Békésy (1960)

Neurophysiology of Hearing

THE COCHLEAR MICROPHONIC

Whatever the nature of the method of transformation of energy may be in the ear, the ultimate result must be the instigation of neural activity in the nerve fibers of the auditory system These may be expected to be accompanied by electrical activity One of the first reports of electrical activity in the auditory mechanism was that of Wever and Bray (1930) Recording from the round window of the cat by means of a gross electrode, they identified an electrical response which followed the temporal characteristics of the driving auditory stimulus with remarkable accuracy It was at first thought that this was the electrical concomitant of neural activity. Subsequent findings to the effect that this activity had virtually zero latency, was relatively less susceptible to the effects of reduced oxygen than neural tissue, and the fact that it "followed" temporal rates of change quite in excess of those which nerve fibers may be expected to follow, all suggest that this *cochlear microphonic* is more probably a by-product of the stimulation of the auditory mechanism, or an *epiphenomenon* associated with the transformation of the mechanical energy of the driving signal into neural activity

Actually, physical disturbance or vibration of organic tissue may lead to voltage changes in a wide variety of circumstances This is true of many of the tissues within the ear The primary source of the cochlear microphonic reported by Wever and Bray, however, the "first order cochlear microphonic" as von Békesy terms it, probably arises in the hair cells A variety of experiments have been performed which demonstrate that the polarity of the cochlear microphonic potential, or its phase, is reversed as a small electrode from which it is recorded passes through the region of the organ of Corti associated with the interface between the hair cells and the tectorial membrane (von Békesy, 1960, Konishi and Yasuno, 1963)

As the amplitude of a driving signal is increased, the amplitude of the cochlear microphonic also increases and it continues to follow the form of the driving signal quite accurately With still further increase in driving amplitude the amplitude of the cochlear microphonic begins to level out and may even reduce Even so, the wave form is not distorted This is a very unusual characteristic for any simple physical system It suggests that the cochlear microphonic as recorded is certainly not reflective of the activity of single units within the system, i e , single hair cells, but rather represents the summated activity of many units Fidelity of following of the wave form of the driving signal, even when amplitude begins to saturate, may reflect a low pass filtering characteristic of the overall system which serves to filter out nonlinearities imposed by the system (Whitfield, 1967)

A number of experiments have demonstrated that it is possible to record cochlear microphonics representative of a much narrower extent of the basilar membrane than the original signals recorded by Wever and Bray This is accomplished by locating electrodes on either side of the organ of Corti, one in the scala vestibuli and one in the scala tympani Recording with a differential amplifier, it is possible to exclude other signals and detect the cochlear microphonic representing as little as 1 or 2 mm of the length of the basilar membrane and its associated tissue (Tasaki, et al., 1952, Tasaki and Fernandez, 1952) The cochlear microphonic has proven a useful indicator for the determination of what portions of the basilar membrane are responding to what frequencies of the driving stimulus (Davis et al , 1953) The voltage of the cochlear microphonic in the first turn follows the velocity of the movement of the stapes and not its absolute position It thus represents the first derivative of the driving force and this serves to

explain the fact that there is an initial 90° phase lag of the cochlear microphonic with respect to the driving signal (Teas, 1970)

The mechanism of action of the hair cell remains a mystery Recently, Harris et al (1970) have demonstrated that the hair cells of the lateral line organ of the mud puppy respond with a graded potential to deflection of their cilia It has been suggested that the hair cells may be important by reason of the fact that they provide additional effective cell surface across which an electrostatic potential may appear, and that it is the movement of large charged molecules which gives rise to microphonic potentials (Dohlman, 1960)

TEMPORARY THRESHOLD SHIFT

It is not surprising that with prolonged exposure of the ear to a relatively loud sound of a given frequency there is an elevation of the auditory threshold for some minutes following exposure It is curious, however, that this temporary threshold shift (TTS) is greater for frequencies higher than that of the initial exposure by ½ to 1 octave The frequencies affected to a greater extent cause motion of the basilar membrane which is confined to a region closer to the base of the membrane Whatever the process, one might suppose that it would be greatest in the region of maximum amplitude of motion of the membrane as caused by the "fatiguing" frequency The leading edge of the envelope of the traveling wave of the frequency most affected corresponds more nearly with the point of maximum amplitude of the envelope of the traveling wave for the "fatiguing" frequency This suggests the possibility, if the effect is greatest at the point of maximum amplitude, that the portion of the membrane which is most significant for the detection of a given frequency of stimulation lies somewhere apical to the point of maximum amplitude This would be in a region in which von Békésy has suggested that the direction of shear of the hair cells is longitudinal It is difficult to resolve with von Bekésy's contention (1960) that the greatest cochlear microphonic arises basal to the point of maximum movement for a given frequency of stimulation There has been one experiment reported which would appear to provide an electrophysiological analog for the TTS frequency shift Stevens, Davis, and Lurie (1935) reported that damage to an isolated region of the cochlea of the guinea pig suppressed cochlear microphonics for frequencies higher than that which gave rise to maximum amplitude of motion of the basilar membrane in the region of the damage Subsequent studies of the correspondence between motion of the membrane and damage tend to show that damage does in fact occur in the region of maximum amplitude of motion (Davis et al , 1953)

HAIR CELL FUNCTIONS

Possible divisions of the functions of the outer and inner hair cells have long been a matter of interest It has been suggested that the inner hair cells, smaller in number, may be analogous with the cones of the retina and responsible for fine discriminations at higher sound pressure levels, while the outer hair cells are more involved in the detection of near threshold events It is recognized that there may be differences in the nature of neural connections to inner and outer hair cells Outer hair cells have been reported more often to be connected to fibers which enter the organ of Corti and then travel longitudinally along it before termination Inner hair cells are said to be connected with fibers that emerge from the nerve bundle and terminate directly on hair cells without traveling along the cochlear mechanism Suggestions that connections from inner hair cells may be more direct and that many outer hair cells may be connected to a single pathway have not been substantiated by recent electronmicroscopy (Whitfield, 1967)

Inner hair cells seem to be better protected from physical damage than outer hair cells (von Békésy, 1960) but would appear to be more sensitive to reduced oxygen levels than are the outer hair cells If indeed the inner hair cells prove to be activated by longitudinal bending near the apical front of the traveling wave envelope and are primarily responsible for frequency discrimination, while the outer hair cells are activated primarily by lateral bending in the region of maximum amplitude of excursion of the membrane and basal to this point, then the frequency shift which accompanies a temporary threshold shift could readily be explained There are too many conflicting items of evidence for this to be accepted as a firm conclusion at this juncture, however

It has been estimated that when the ear is stimulated by a 40-db tone at 1000 cycles, approximately 3000 hair cells may be activated With a change in frequency by a just discriminable amount, it is probable that less than 2%

of these hair cells are no longer activated and have been supplanted by others which were formerly inactive A great deal of evidence suggests that it is the overall pattern of stimulation of many receptors rather than the activity of just a few which is involved in even the relatively simple discriminations of pure frequencies from one another

ELECTRICAL POTENTIALS

Following discovery of the cochlear microphonic and the recognition that it could not represent an index of neural activity, the search for other electrical evidences of neural action continued

Summation Potential. Among other electrical indicators that have been identified is the summation potential The cochlear microphonic potential is accompanied by a shift in the direct current (DC) level of the base line There is an increase of voltage in the scala tympani with respect to the scala media This may be reversed by an increase of pressure in the scala tympani Whitfield (1967) has pointed out that the summating potential cannot be a DC component introduced by the nonlinearity of the response of the cochlear microphonic, because the amplitude of the summation potential may differ along the basilar membrane A peculiarity of the summation potential is the fact that its amplitude may increase with a decrease in temperature This is the opposite of what would be expected if it were of chemical origin Whitfield has concluded that it is a mechanical effect which may be attributed to asymmetry of the basilar membrane It has been shown that, just as is so for the cochlear microphonic, the polarity of the summation potential changes when the recording electrode penetrates the reticular lamina or region of the interface between hair cells and the tectorial membrane Its source is thus localized in the region of the hair cells (Konishi and Yasuno, 1963)

A recent study (Dallos, Schoeny, and Cheatham, 1970) has shown that the summating potential, measured between the scala tympanie and the scala vestibuli, varies in form along the cochlea with the frequency and the intensity of stimulation Of particular interest is the fact that the average of the potentials in each of these canals apparently does not reflect just remote activity with a cancellation of the local cochlear microphonic potential A negative constituent probably represents remote activity, but the positive constituent of the average is dominant in the region of maximum excitation and may represent a hyperpolarization effect in a segment of the cochlea associated with the mechanism of receptor excitation It is also of interest in relation to our discussion of temporary threshold shift above, that the peak of the summation potential, whether it is based on the difference between the scala tympani and the scala vestibuli or the average potential for those two regions, corresponds with the distal end of the traveling wave envelope rather than the point of maximum mechanical vibration of the cochlea for a particular frequency of stimulation

Regional Potentials. Relatively steady potentials may be recorded in various regions of the cochlear mechanism When referred to the perilymph, the voltage recorded anywhere within the organ of Corti is approximately −20 to −80 mv The endolymph of the medial canal, on the other hand, is at a voltage of +80 mv There is thus a voltage difference between the endolymph of the medial canal and the organ of Corti of the order of 140 mv The high voltage of the endolymph is somewhat unusual The voltage of the endolymph in the vestibular end organs is only 1 to 5 mv The endolymph of the scala media is highest in the region of the stria vascularis and has been attributed to metabolic activity of that tissue This voltage has been shown to depend upon an adequate oxygen supply and has also been demonstrated to fluctuate with motion of the basilar membrane This endocochlear or endolymphatic potential increases with downward motion as the stapes presses the oval window in and decreases with upward motion of the membrane as the oval window is drawn out (von Békésy, 1952, Tasaki, Davis, and Eldridge, 1954) The endocochlear potential is not analogous to a typical cell membrane potential because it is uninfluenced by changes in the potassium concentration of the scala tympani It would be difficult to explain in these terms in any case when one compares it to the relatively low voltage of the virtually identical endolymph in the labyrinthine mechanism It can be demonstrated fairly simply that the endolymph with its high potassium concentration has no access to the receptors within the organ of Corti If the potassium concentration in the scala tympani is increased by a small

amount, the cochlear microphonic is rapidly attenuated This suggests that the basilar membrane is permeable to potassium, but that the boundary separating the organ of Corti from the endolymphatic space is not A positive potential is always found in the region of the stria vascularis even when the endolymph is removed Thus, the high potential in this region is attributed to activity within the stria vascularis rather than to chemical action directly involving the endolymph

The Neural Action Potential With all of the electrical activity in the ear, it has been difficult with gross electrodes to isolate the action potential of the nerve fibers themselves One method of accomplishing this has been to add signals from two electrodes, one on each side of the organ of Corti The purpose is to cancel out the cochlear microphonic and reveal the whole nerve action potential Another procedure has been to use a relatively short stimulus such as a click With this procedure, in a recording from the round window the cochlear microphonic can be seen immediately, while the action potential is observed after a short latency (Fig 8 16) The latency of the action potential is correlated with frequency and the region of maximum activity along the cochlear partition, the higher the frequency the shorter the latency One interesting application of measurements of the action potential is its use as an index of the possible effect of efferent signals to the cochlear mechanism in the suppression of response to auditory stimuli

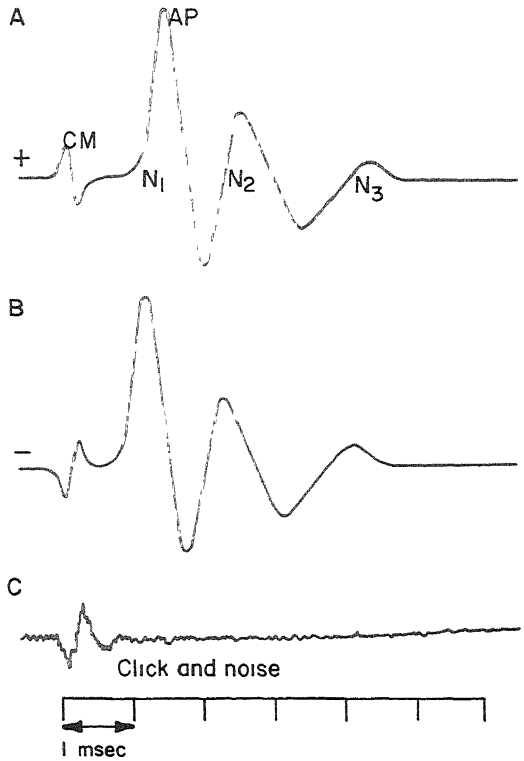


Fig 8 16 The electrical responses to clicks as recorded from the round window of the cochlea of the cat CM cochlear microphonic potential AP action potential of the auditory nerve The three components of the action potential are represented by N_1 N_2 and N_3 A reversal of polarity of the stimulus between A at the top of the figure and B results in an inversion of the cochlear microphonic but no change in the action potential except for a slight shift in latency At C the action potential is completely masked by white noise while the cochlear microphonic remains unchanged

INNERVATION OF THE COCHLEAR MECHANISM

It is estimated that approximately 30,000 auditory nerve fibers and ganglion cells (Rasmussen, 1940) subserve the total of approximately 25,000 hair cells in man (Guild, 1932) The auditory nerve contains both afferent and efferent fibers which terminate on the hair cells All of these reach the peripheral mechanism via the auditory nerve, through the core of the modiolus, through tiny holes, the habenula perforata, and then over the basilar membrane where they are unmyelinated. The work of Fernandez (1951) permits the division of these fibers into several types Radial fibers extend out from the modiolus and terminate primarily on inner hair cells Some of them penetrate the tunnel of Corti to terminate on external hair cells Longitudinal fibers extend out radially and then turn to travel as much as ⅓ to ½ turn along the cochlea in the basal direction These terminate primarily on outer hair cells but may also terminate on inner hair cells The radial fibers are attached to a relatively small number of adjacent cells in the case of both outer and inner hair cells The longitudinal fibers on the other hand, connect with many outer hair cells and may project terminations to cells in different rows along their course Just as the fibers diverge and terminate on many cells, so do processes from different fibers converge on single cells (Figs 8 17 and 8 18) The nerve endings which cluster about the hair cells may be divided into two different types the smaller ones are 0 5 to 1 0 μ in diameter with few vesicles,

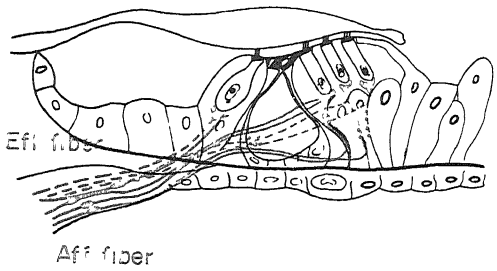


Fig 8 17 The projections of afferent (Aff) and efferent (Eff) fibers onto outer and inner hair cells of the organ of Corti (Wersall Flock and Lundquist 1965)

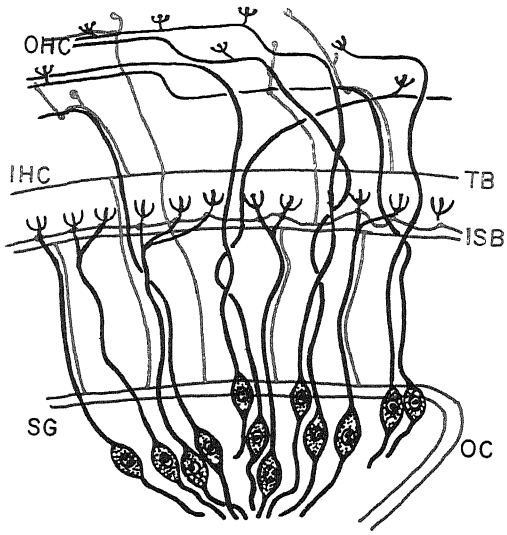


Fig 8 18 Schematic illustration of the innervation of cochlear hair cells Afferent fibers are in solid black lines efferent fibers are represented by brown lines OHC outer hair cells IHC inner hair cells SG spiral ganglion OC olivocochlear bundle ISB internal spiral bundle ESB, external spiral bundle TB tunnel bundle (Smith and Rasmussen 1963)

while the larger type may be as large as 5 μ long and 1 to 3 μ in diameter and is densely filled with vesicles Electronmicroscopy reveals presynaptic bars under the plasma membrane of the hair cells where it is in contact with the smaller endings The membrane adjacent to the larger endings is characteristic of a postsynaptic relation

The smaller endings are believed to be afferent, while the larger ones are believed to represent the terminations of efferent fibers from the olivocochlear bundle Rasmussen (1960) has identified about 500 of these efferent fibers in the cat and observed terminations on both inner and outer hair cells The relative courses of cochlear and vestibular efferent fibers are illustrated in Figure 8 19 Bipolar cell bodies of the afferent neurons are located in the spiral ganglion, within Rosenthal's canal of the modiolus Central to the ganglion they form the cochlear nerve trunk, fibers of which terminate on the dorsal and ventral cochlear nuclei

RESPONSE CHARACTERISTICS OF SINGLE UNITS

It is of considerable interest to know the way in which individual nerve cells of the auditory nerve bundle respond to various frequencies of stimulation of the ear The first experiment to report the responses of single units was that of Galambos and Davis (1943) Using microelectrodes, they obtained responses from single fibers and explored the way in which the discharge frequency varied as a function of the frequency of stimulation From such records it was possible for them to plot response area functions for the individual units (Fig 8 20) Units were found to have a single "best" frequency at which they would respond with minimum sound pressure With increase in the sound power of the stimulating source, these units showed a widening range of frequency over which they would respond Characteristically, when plotted on a logarithmic scale, the response area function showed a steeper front toward the higher frequencies This suggests that the increased range of frequencies with increased sound pressure was gained largely in the direction of lower frequencies Some correction must be made for the logarithmic scale, however Nonetheless, one might expect a greater increase in the frequency of response toward reduced frequencies by reason of the fact that increased amplitude will result in an extension of the motion of the basilar membrane toward the apex, the direction which will bring in those hair cells which presumably are responding to low frequencies

Subsequent findings (Galambos and Davis, 1944) of inhibitory processes placed in question the original conclusion that recordings had been made from the primary auditory nerve fibers The nature of inhibitory processes is still not entirely clear, however, and the demonstration of inhibition did not necessarily indicate that the source of original recordings was secondary cells of the cochlear nucleus rather than the fibers of the auditory nerve

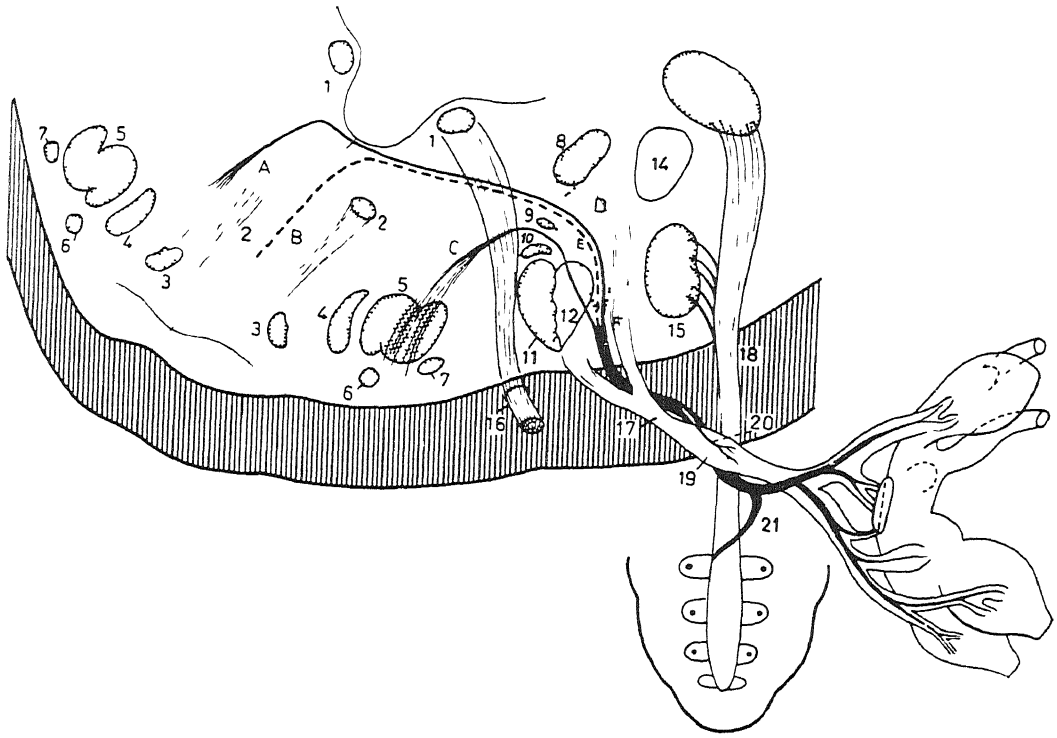


Fig 8 19 Cochlear and vestibular efferent systems A crossed cochlear efferent bundle B uncrossed reticulocochlear and reticulovestibular bundle C uncrossed cochlear efferent bundle D and E dorsal and ventral uncrossed vestibular efferent bundles F bundle containing all of the efferent cochlear and vestibular fibers 1 facial genu 2 abducens 3 trapezoid nucleus 4 medial superior olivary nucleus 5 lateral superior olivary nucleus 6 medial preolivary nucleus 7 lateral preolivary nucleus 8 lateral vestibular nucleus 9 interposed vestibular nucleus 10 inferior vestibular nucleus 11 nucleus of descending root of trigeminal 12 descending root of trigeminal 13 dorsal cochlear nucleus 14 inferior cerebellar peduncle 15 ventral cochlear nucleus 16 facial nerve 17 vestibular division of N VIII with superior and inferior roots 18 cochlear division of N VIII 19 ganglion of Scarpa 20 efferent fibers to Scarpa s ganglion 21 vestibulocochlear anastomosis of Oort (Rossi 1965)

Subsequently, large numbers of recordings have been made from single fibers in the auditory nerve in the cochlear nucleus by Tasaki (1954), Katsuki and his associates (Katsuki, Sumi, Uchiyama, and Watanabe, 1958, Katsuki, Suga, and Kanno, 1962) and by Kiang (1966) The response area curves plotted against a logarithmic frequency scale were found to be both symmetrical and asymmetrical with a steeper front toward higher frequencies in the latter case The symmetrical units were less common than asymmetrical ones It has not been possible to demonstrate specific associations of symmetrical units with the different receptors of the cochlea Response areas appear broader for lower frequency units

The finding of response areas with characteristic best frequencies distributed over the entire range of frequency response for the ear studied is a powerful argument for the encoding of frequency information in the auditory system in terms of the individual fibers or patterns of fibers excited It has, therefore, been of interest to examine the frequency response characteristics of units higher in the auditory nervous system in order to ascertain the extent to which this type of coding is retained, as information is transmitted to the primary auditory regions of the cortex

NEURAL ANATOMY OF THE AUDITORY SYSTEM

Afferent fibers, which mediate the transmission of responses of the receptor cells within the organ of Corti after entering the medulla in the cochlear nerve, divide and send branches to the dorsal and ventral cochlear nuclei (Fig 8 21) A number of second order neurons pass from these to the superior olivary

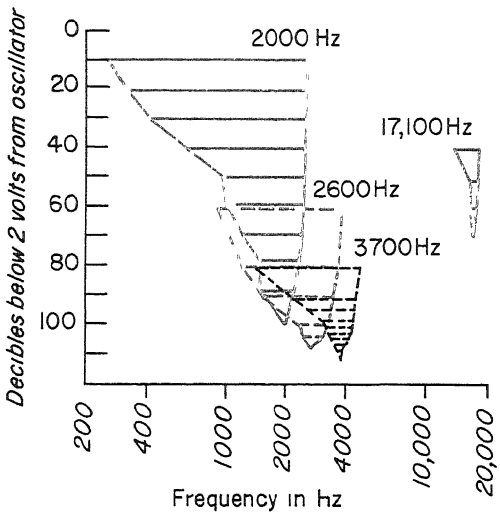


Fig 8 20 Response areas of four neurons in the cochlear nucleus of the cat Each neuron has a specific frequency at which a minimal acoustic energy is required to excite it Each responds also to progressively lower frequencies at higher intensities There is a slight extension of the range of response into higher frequencies but the cut-off is much sharper on the high frequency side Some distortion of the shape of these cut-off functions as a result of the logarithmic frequency plot must be taken into account (Galambos and Davis 1943)

nuclei on both sides from *each* side of the medulla Other fibers ascend to the lateral lemniscus directly, terminating in its dorsal nucleus or extending on to the inferior colliculus At each of these levels there appears to be crossing by fibers from one side to the other There is thus representation on each side of the brain of fibers carrying signals which have originated in each ear Fibers from the inferior colliculus ascend to the medial geniculate nucleus ipsilaterally From this station in the medulla, fibers pass to the primary auditory cortex via the internal capsule In addition to the direct pathway described, there are also additional reflex connections with various cranial motor nuclei The fact that a cat with complete transection of the classic auditory pathways at the level of the midbrain can be aroused by loud sounds suggests a secondary route to the cortex, perhaps via the reticular formation of the midbrain

Laterality of the Auditory System. Although anatomical pathways are available to both sides of the brain from each ear, there is some evidence that crossed pathways play the dominant role (Rosenzweig, 1951, Kimura, 1961b) Thus, the right ear provides better access to the left hemisphere and the left ear provides better access to the right hemisphere In a variety of experiments, evidence has been found for selective hemispheric processing of auditory inputs depending on their nature Language related sounds are usually processed by the left hemisphere and other sounds are processed predominantly by the right hemisphere Evidence for this comes from simultaneous presentation of different sounds to the two ears In the case of speech sounds, such as digits or consonant-vowel combinations, the sound presented to the right ear is retained more frequently when competing sounds are presented to the two ears (Kimura, 1961a, Broadbent and Gregory, 1964) The reverse is true for nonspeech sounds, such as music (Kimura, 1964) and sustained vowel sounds or tonal sounds (Shankweiler and Studdert-Kennedy, 1967) These findings are supported by studies of patients with damage to the right and left temporal lobes Left temporal lesions are associated with poorer performance in the recognition of digits than right temporal lobe lesions Performance in the reporting of digits presented to either ear is affected by left temporal lobectomy, but not by right temporal lobectomy (Kimura, 1961a) In some individuals, the right hemisphere may be dominant for language, in which case the left ear provides a more ready access for speech sounds (Kimura, 1961b)

Efferent Pathways. In addition to the ascending pathways, descending pathways from the cortex have been described (Desmedt, 1962) Fibers which terminate on both the cells of origin of the olivocochlear bundles and on the neurons of the ventral cochlear nucleus have been identified The olivocochlear bundle originating in the superior olivary nucleus is an efferent pathway from which the efferent fibers terminating on hair cells arise The identity of the efferent fibers and the afferent fibers has been ascertained by sectioning of the olivocochlear bundle, in which case efferent fibers degenerate If the acoustic nerve is sectioned, afferent fibers degenerate These procedures permit the identification of which nerve terminals are which in the region of the hair cells (Engstrom, Ades, and Hawkins, 1965a) The precise function of the efferent pathways is not clear, but an inhibitory role has been assigned to them (Galambos, 1956)

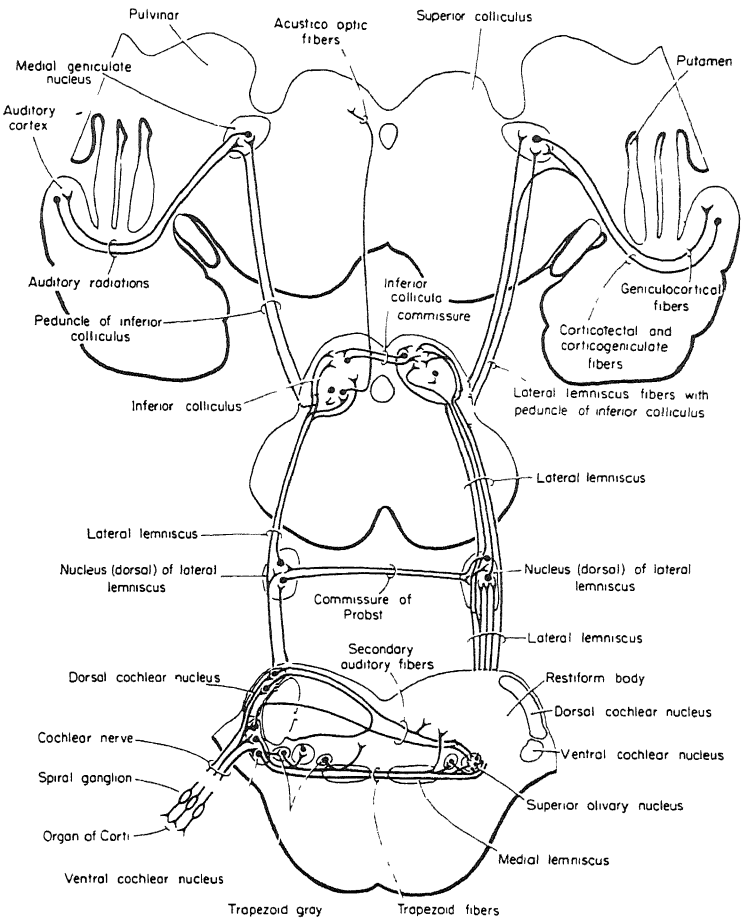


Fig 8 21 Major central pathways of the auditory system (Crosby Humphrey and Lauer 1962)

VARIATION OF FUNCTION AT DIFFERENT NEURAL LEVELS

Response areas have been studied in all parts of the auditory nervous system (Fig 8 22) It is evident that the frequency coding of single units is maintained up to the level of the inferior colliculus At this point there are a number of relatively narrow response areas with best frequencies distributed over the entire auditory spectrum Katsuki has argued (1961) that this type of coding is maintained up to the level of the medial geniculate body, but the evidence does not appear to be as good at this level, even considering Katsuki's own work Whitfield argues (1967) that tonotopic organization is of little consequence at levels higher than the inferior colliculus On the other hand, Erulkar has argued (Erulkar, Nelson, and Bryan, 1968) that some tonotopic organization persists up to the level of the cortex with a variation in selectivity depending upon the cortical layer from which recordings are made

Surveying all of the evidence, it is clear that the nature of response of single neural units changes from the auditory nerve up to higher centers with good frequency coding up to the level of the inferior colliculus, and relatively poor frequency coding beyond this point toward the auditory cortex In addition, there is a relative decrease in the correlation between the intensity of the stimulus and the frequency of discharge of single units as one ascends in the system There would appear to be a significant change in the nature of coding from the level of the inferior colliculus up to the cortex This is discussed further below

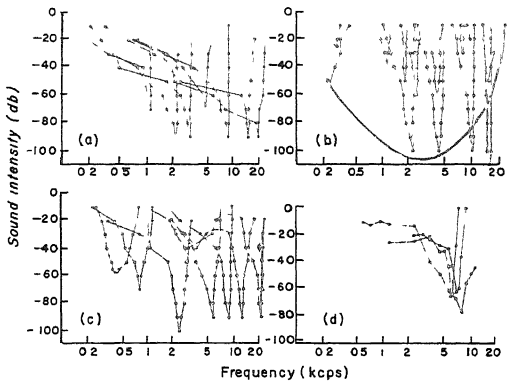


Fig 8 22 Response areas of single neurons obtained from *a*) cochlear nerve *b*) inferior colliculus *c*) trapezoid body and *d*) medial geniculate body of the cat (Katsuki 1961)

INHIBITORY PROCESSES

Inhibition is recognized as playing an important role in all sensory systems Several different kinds of inhibition have been observed in the auditory system Some of these originate in the peripheral mechanism itself and others may have their origin at higher levels of the auditory nervous system Any reduction in the level of activity of a single fiber or unit which accompanies stimulation is defined as inhibition This may result by the presentation of a single tone which serves to reduce the spontaneous firing rate of a unit not otherwise stimulated (Ruppert et al , 1963, Allanson and Whitfield, 1955) Another way of demonstrating inhibition may be to reduce the activity resulting from the effects of continuous tone stimulation by the presentation of a second stimulus (Frishkopf and Goldstein, 1963) Under these circumstances, the inhibitory signal may be one to which the unit would show no response at all, or it may be one to which there would be a positive response in the absence of any other input stimulation The results of a variety of investigations do not provide good agreement on the precise nature of these inhibitory processes No inhibitory effects below 1000 Hz. have been reported by Nomoto et al (1964), while Sachs and Kiang (1968) report two-tone inhibition effects at frequencies as low as 200 Hz The latter investigators report that, when tested appropriately, all fibers in the auditory nerve appear to show two-tone inhibition On the other hand, they found no evidence for suppression of spontaneous activity in any of the units which were tested with a single tone (Fig 8 23) Here, positive response areas are seen to be bounded by inhibitory regions which intrude slightly into the positive response area Within the inhibitory regions, stimulation suppresses activity elicited by an exciting stimulus

The function of these inhibitory processes may be to provide a sharpening of the frequency selectivity of the system (Greenwood and Maruyama, 1965) Erulkar, Gerstein, and Butler (1966) report more evidence for inhibitory processes up to the level of the inferior colliculus than within the cortex Their work is based on intracellular recording of the inhibitory postsynaptic potential

Most of the inhibition discussed above probably arises out of neural interactions among peripheral neural elements It has been suggested, however, that two tone inhibition could result from mechanical interaction in the basilar membrane (Simmons, 1970)

There is another mechanism for inhibition quite different from those described above This results from the activity of efferent fibers of the olivocochlear bundle, arising in the superior olivary nucleus and innervating, contralaterally, hair cells of the organ of Corti Galambos (1956) has demonstrated that stimulation of the olivocochlear bundle via the floor of the fourth ventricle may reduce the whole nerve response by as much as would a 15- 20-db drop in stimulus intensity At the same time, such stimulation may be accompanied by an enhancement of the cochlear microphonic potential (Desmedt, 1962, Fex, 1962)

POSSIBLE BINAURAL MECHANISMS

Some units respond to stimulation of the two ears with differences in latency The magnitude of the latency difference is of the appropriate order of magnitude for the difference in time of arrival of a sound pressure wave at the two ears when the source of the sound is located off to one side of the listener The presumption is that these cells would be maximally excited by sound originating in just such a location For such a unit, a sound on the opposite side of the body would result in the units being relatively refractory at the time of excitation of the distal ear Such units could thus serve in binaural localization (Erulkar et al , 1968) Erulkar has also found units for which the nature of the frequency response area differs between the two ears What function such units may serve is unclear

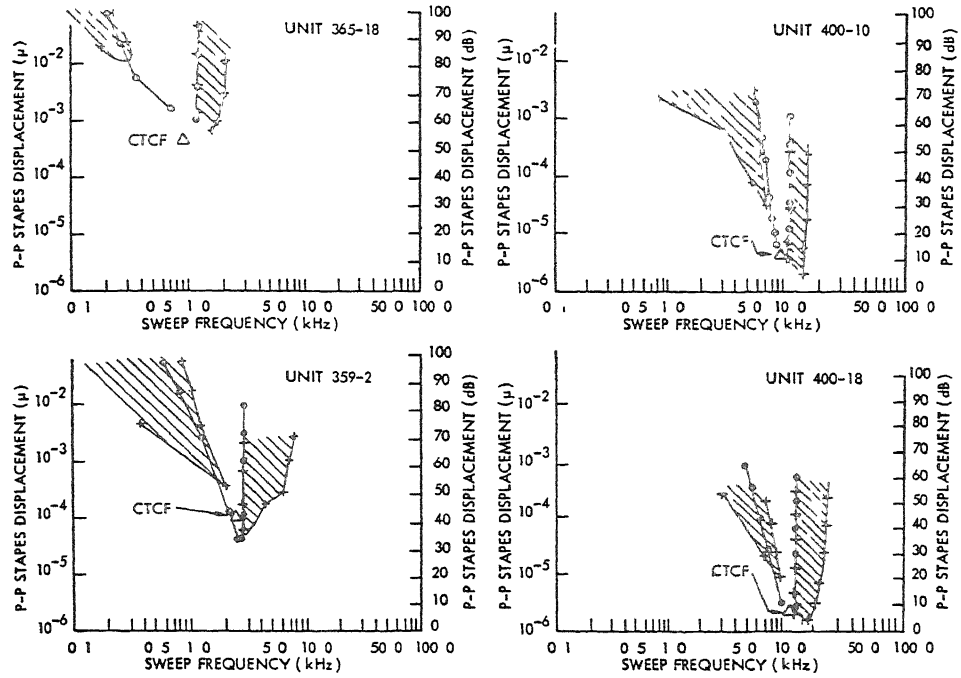


Fig 8 23 Response and inhibitory areas for four fibers in the cat auditory nerve The characteristic frequency (CTCF) is marked by a triangle and represents that frequency to which the unit is most sensitive The boundary of the response area marked by dots is defined by the amplitude of a sweep frequency which results in an increase in rate by 20% over the level of spontaneous activity Ordinate axes represent peak-to-peak stapes displacement in micra and in decibels The hatched areas represent frequencies and amplitudes in which the sweep frequency reduces the response rate by 20% or more below the rate obtained with the characteristic frequency alone These inhibitory areas are defined by presenting the critical frequency continuously along with the sweep frequency the latter varying exponentially in time between two frequency values which differ by one or two decades (Sachs and Kiang 1968)

THEORIES OF AUDITORY INFORMATION PROCESSING

The adaptive role of the auditory mechanism in man or animals is to provide information concerning the environment Sound is an important dimension for warning, searching, identifying, and signalling As a medium of communication, it is at its most complex in human speech To understand the mode of function of any sensory system, attempts are often made to determine ways in which it permits the encoding of simple physical dimensions of the stimulus for transmission to the brain for analysis and interpretation A weakness of this approach is that it often begins with the selection of certain stimulus dimensions, considered in their simplest form, and the evaluation of how these influence the nervous system when varied parametrically In audition, this has lead to great concern with the encoding of frequency information As suggested earlier, this may not be unreasonable in view of the fact that complex sounds can be synthesized by the combination of appropriate sinusoidal components It is possible however, that such endeavors have been overemphasized Nonetheless, work on the processing of frequency information has been useful It has demonstrated a relation between the response of specific neural units and the frequency of the stimulus and, at low frequencies at least, some relation between temporal periodicities of the stimulus and the response

These relations are clearly less pronounced at higher auditory centers In human voice communication, relatively little information is transferred on the basis of simple frequency coding Speech sounds, for example, have a recognizable identity by reason of the relative temporal variations among their frequency components The actual frequencies involved are relatively unimportant compared to the relations among them Two voices, representing different frequency ranges, can com-

municate the same word with the same meaning with no danger of misinterpretation The identity of the word clearly depends upon relations which are independent of specific frequencies (Garrett and Fodor, 1968) The identity of the word and its recognizability are also quite independent of its loudness, provided that loudness is above a certain minimum level for the conditions of background noise which prevail

What may serve best for analysis then, is a system which includes elements that will respond to selected temporal variations among components of a stimulus where relative frequencies are of greater importance than absolute values Many of the cells of the auditory cortex seem to respond in just this way While peripherally there is clear tonotopic organization of the system, with specific fibers responding maximally to specific frequencies of stimulation and specific cells of the cochlear nucleus in the inferior colliculus doing likewise along with physical arrangement of units in locations which show some relation to their frequency response, this kind of organization is far less pronounced in higher centers, particularly in the cortex With ascent in the auditory nervous system beyond the inferior colliculus, there is a decrease in the relation between the frequency of neural activity and the intensity of the stimulus, and a decrease in the extent to which single units are associated uniquely with a "best frequency" Some tonotopic organization persists in the cortex (Hind, 1960), but it is greatly reduced At the level of the medial geniculate nucleus, many cells may respond only to the onset of stimulation or to its offset At the level of the auditory cortex investigators have been frustrated by an inability to find any appropriate stimulus for a number of units which can be detected by their spontaneous discharge In anaesthetized animals, Erulkar, Rose, and Davies (1956) found 34% of the units studied totally unresponsive In unanaesthetized animals, Galambos (1960) found 18% unresponsive He was able to characterize the response of approximately 70% of the cells studied in terms of the optimum frequency of pure tone stimulation, but other investigators have not found as high a percentage as this Whitfield and Evans (1965) report that 21% of cortical units respond only to "odd" sounds such as the jingling of keys, and that a significant percentage cannot be stimulated at all Units which respond to change in frequency or amplitude have been found, and some of these are sensitive to the direction and rate of change of frequency (Erulkar, Nelson, and Bryan, 1968) Some units of this sort have been reported in cochlear nucleus cells, but they are much more common in the cortex Such units must be a focus of converging fibers from peripheral regions which represent a range of frequencies, and possess temporal response characteristics which vary systematically with their frequency response characteristic It is possible that the nature of cortical connections is determined during development, after birth, and is influenced by the nature and relevance of the stimuli to which the organism is exposed Appropriate stimulation for many of the cortical units that have been detected by reason of spontaneous discharge has simply not been found It would be impossible to explore all of the logical stimulus possibilities in the laboratory, there are too many of these (Erulkar et al , 1968)

Another characteristic of the nervous system which renders it difficult, if not impossible, to arrive at a definitive statement as to how it works is its great flexibility Animals which have been trained to make discriminations based on frequency of auditory stimulation show a disruption of their performance after ablation of the auditory cortex, yet they are able to relearn such discriminations following ablation Thus, such discriminations can be mediated by other, probably lower, centers (Neff, 1961) Initial disruption of the performance argues for some cortical role under normal conditions, and it is possible that discriminations learned after ablation are based on neural cuing systems which never operate in a normal animal. It is of course possible that the initial disruption is a sequel to the operative procedure and that pitch discriminations are characteristically made at lower levels in the nervous system

Cortical ablation does disrupt certain kinds of performance irrevocably Any complex temporal discriminations, such as those based on the pattern of repetition of two tones, cannot be relearned in the absence of the cortex It therefore is reasonable to conclude that the most important discriminations made by auditory mechanisms, such as the recognition of speech sounds, do in fact depend upon cortical function Some understanding of how this

complex system may work has just begun to develop There are some interesting analogies between the auditory mechanism and the visual mechanism

Basic Auditory Functions and Hearing Impairment

PSYCHOLOGICAL STUDIES OF HEARING

A variety of properties of the auditory mechanism of man has been studied by psychophysical procedures (Licklider, 1951, Tobias, 1970) These include studies of the way in which characteristics of the stimulus must be varied for a listener to make some sort of discrimination for a given set of background stimulus conditions, or for a listener to make a judgment of equality for some specific dimension between two different stimuli An example of the latter might be the loudness of two tones of different frequency The results of such experimentation provide us with valuable information on the nature of the function of the auditory mechanism, and they are also of practical concern for the testing of this mechanism in order to discover evidences of impairment

Work of this kind has been extended in another class of investigation in which listeners are asked to make magnitude estimations or to adjust stimuli for ratio changes with respect to some dimension For example, a listener might be asked to set the intensity control of an oscillator such that a given tone has a loudness twice as great or half as great as that of a tone heard in an interval preceding the time at which the setting is to be made In general, studies based on a simple discrimination task or on an equality judgment provide more reliable results, but all of the results of investigations of this sort are of interest and provide a basis for further insights into the nature of the function of the auditory system

Absolute Threshold. The smallest amount of sound power that the ear can detect in any given frequency is necessarily a statistical concept Attempts to measure such an amount produce variable results which are influenced by differences in the subjects of the test, differences from time to time in a given subject, and a variety of procedural factors Nonetheless, minimum sound power levels must be selected which can serve as a standard representative of a large population of normal listeners. Two functions which are representative of the minimum threshold in sound pressure are illustrated in Figure 8 24 including the recent international audiometric zero curve (ISO, 1964) which serves as a current standard

Differential Thresholds. Differential thresholds, at least for the discrimination of intensity differences, may be considered as an extension of the absolute threshold determination when the background is changed from one of quiet to one with a controlled continuous sound level In many sensory modalities it has been demonstrated that there is a rough relation between the level of intensity increment necessary for discrimination and the intensity level of the background against which the discrimination is to be made This relation is known as the Weber-Fechner Law, but is really only an approximate description over a limited range of the variables which may be studied In audition, the intensity increment required for discrimination at background levels near threshold is relatively large but declines in relative amplitude up to a background level of approximately 50 db above threshold The situation is complicated by the fact that the size of the increment also varies with the frequency of the sound

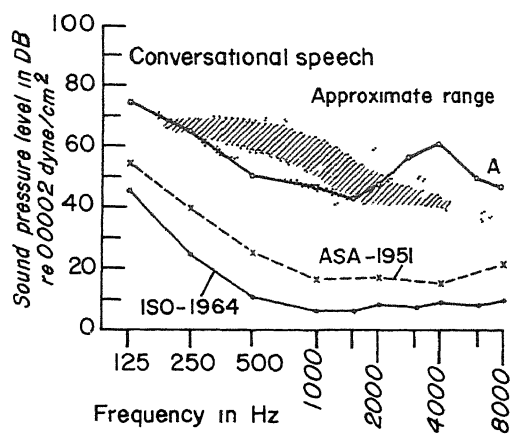


Fig 8 24 The newly adopted International Audiometric Zero (ISO-64) curve compared with the American Standards Association (ASA) 1951 curve The ISO curve represents a better approximation of the thresholds of normal young adults than does the ASA 1951 curve The shaded area indicates the pressure-frequency region in which most conversational speech sounds occur Curve A represents thresholds typical of a case of moderate hearing impairment (Davis 1965)

When the change in frequency of a sound necessary for discrimination of the fact that frequency has been changed is studied as a function of frequency, there is an approximately constant ratio between the frequency increment and the reference frequency above 1000 Hz The absolute increment in frequency is roughly constant at 2 or 3 Hz from 60 to 1000 Hz for sound pressure levels above 30 db A three-dimensional plot of this latter relation is illustrated in Figure 8 25, with sensation level in decibels as the third parameter

The nature of the frequency response of the human ear can be examined from the level of threshold for detection of sound at a given frequency up to the maximum comfortable level of sound stimulation by obtaining data on the matching of loudness at a variety of levels over the entire frequency spectrum These relations are illustrated in Figure 8 26 Each curve is labeled in accordance with the sound pressure level in decibels at 1000 Hz It is clear that there is considerable flattening of the frequency response with increase in intensity level and hence, loudness This is particularly true at the low frequency end of the spectrum The rate of growth of loudness with increased intensity is thus much greater at very low frequencies The same loudness increase occurs over a much curtailed range of intensities This is also true to a lesser extent at the highest frequencies It is more difficult to obtain data at these high frequency levels for a large number of subjects It is clear that no simple metric for the addition of loudnesses of different frequencies is practical It is impractical not just because of the nature of the data displayed in Figure 8 26, but also by reason of the fact that when differing sound components are added they retain their individual identity to a large degree The listener is able to make some discrimination of the individual components of complex sounds In addition, if frequencies added are sufficiently close together they may produce beating phenomena which further complicate the situation

Masking. From our discussion of differential thresholds for sound intensity, it is clear that with elevation of a background sound the increment which can just be discriminated must also be increased The background sound thus serves, to a degree, to mask the increment The phenomenon becomes more complex when one examines the masking effects of various frequencies on a given frequency or the

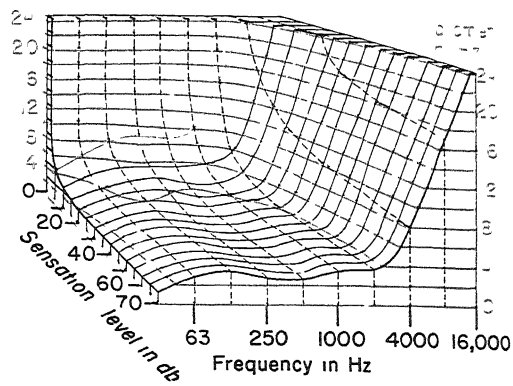


Fig 8 25 Three-dimensional surface showing the standard frequency discrimination threshold as a function of the frequency and the intensity of a standard tone Frequency discrimination is poor at intensity levels near absolute threshold in the rear of the figure and at high frequencies in the extreme right of the figure For frequencies below 1000 Hz at sensation levels above 30 db frequency discrimination requires a change of only about 3 Hz (Licklider 1951)

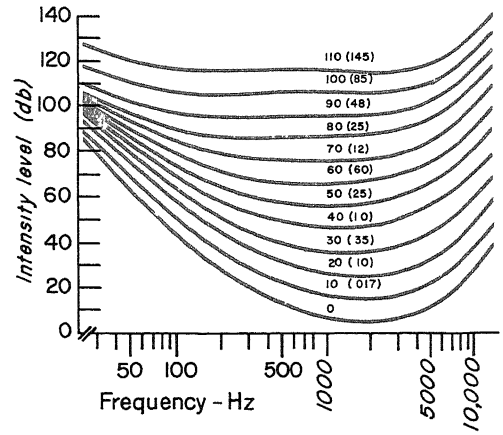


Fig 8 26 Equal-loudness contours showing the relation between intensity level and frequency at various subjectively equated loudness levels The figures 0 to 110 represent loudness level in phons a loudness scale based on equal loudness judgments with the intensity level in decibels of 1000-Hz tone as a reference The lowest curve approximates the normal free-field threshold curve Figures in parentheses show loudness in sones a scale based on subjective ratio estimates The latter scale presumably provides a better indication of subjective magnitudes of loudness (Stevens and Davis 1938)

effects of a wide band of frequencies or noise on a given frequency In general, masking is more effective the closer the frequency of the masking tone to that of the tone being masked, although the precise nature of the relation here may be complicated by the beating

phenomenon Low frequency tones mask high frequency tones more effectively than the converse The latter effect might be expected from our knowledge of the motions of the basilar membrane Interestingly enough, and of some practical significance in the measurement of hearing disability, there is relatively little masking effect when a masking tone is presented to one ear and a test tone to the other This phenomenon is useful in suppressing the better ear when one wishes to test an ear in which unilateral hearing impairment is suspected A fairly broad spectrum noise serves as a useful mask for a wide range of frequencies in these circumstances

Loudness. As illustrated in Figure 8 26, it is possible to make judgments of the equality of loudness, even though the frequency of the tones so equated differs The specification of loudness level is conveniently accomplished by the use of equal loudness contours (Fig 8 26) The intensity level in decibels of a 1000-Hz tone is taken as a reference Loudness level is expressed in *phons* and is equal to the intensity level in decibels of the 1000-Hz tone which matches any other frequency for which the loudness is to be designated

Binaural Cues for Sound Localization A listener is able to estimate the location of origin of a sound moderately well There are several obvious physical cues which may serve to explain this ability in anyone possessed of two normally functioning ears In the first place, a sound originating on one side of the body will reach one ear first and hence, there will be a time delay between stimulation of that ear and the other ear For continuous tones, there will be a difference in the phase relation between the two ears which will depend both on frequency of the stimulating tone and the physical dimensions of the head In addition, the sound reaching the more distant ear may be attenuated The head will serve to cast a sound shadow for the ear on the side away from the point of origin and this will reduce intensity on that side, particularly for sounds of higher frequency

Differences in time or phase are probably of most importance for frequencies below 1400 Hz , while at higher frequencies differences in intensity probably represent the most important cue The higher the frequency, the greater will be the attenuation of sound in the shadow of solid objects past which it is transmitted In careful studies of sound localization, it has been determined that very small displacements of a sound source from the median plane through the head can be discriminated, while much grosser displacements are necessary for the discrimination of a change in point of origin when the sound is directly off to one side or the other Sounds which are directly in the median plane are very difficult to localize as to whether they are in front of, behind, or above the listener Particularly in these circumstances, movements of the head play a very important role in the localization of a sound source It is clear that the changing pattern of intensity between the two ears will be directly opposite for a sound located in front of the head as contrasted with one in back of the head when the head is turned from side to side On the other hand, for sounds which are directly overhead, there will be almost no change at all The absence of any change with head movement is thus a cue to the location of the sound overhead

As anyone who has enjoyed sterophonic music reproduction can testify, there is more to binaural hearing than simple localization Hearing of complex sounds, such as music, the sources of which are distributed in space, is greatly enriched when they are heard with two ears The best stereophonic reproduction is only a gross approximation of this but is nonetheless quite a pleasing phenomenon Two arbitrarily located microphone positions serve to divide the relative balance of the sound which is reproduced in two arbitrarily located loud speakers Optimum effect is achieved when the listener is located in the appropriate relation to the two speakers, and they in turn have a relation which is appropriate for the original method of recording A precise duplication of binaural listening might be achieved by locating two microphones on either side of a dummy head and feeding the sound recorded under such circumstances to a listener via earphones Unfortunately, under these circumstances the changes which one comes to expect with slight movements of the head would not occur.

TYPES OF HEARING IMPAIRMENT

The causes of deafness are divided into five classes (Davis, 1951) In the first of these, there is some interference with the conduction of sound to the neural mechanism of the inner ear. Hearing difficulties resulting from these causes are labeled *conduction deafness* The

second category is that resulting from damage to the cochlear mechanism or the auditory nerve and this is labeled nerve deafness The other three classes include *central deafness*, *diplacusis* or a false sense of pitch which may be caused by edema of the labyrinth, and *tinnitus* associated with ringing in the ears which is usually caused by hypersensitivity of hair cells or their nerves

Conduction Deafness. Conduction deafness may be caused by such a simple problem as the accumulation of wax in the external ear canal It is more significant, however, when it occurs by reason of such factors as the pathological hardening of the tympanic membrane, perforation of that membrane, or a loss of mobility or destruction of part of the ossicular chain A small perforation of the drum is often temporary and may heal, but dysfunction of the ossicular chain requires some kind of intervention for its correction In *otosclerosis*, there is a growth of bone which may tend to freeze the footplate of the stapes such that it is unable to transmit motion to the oval window of the scala vestibuli Initial effects may be manifested by a progressive loss of hearing for the low frequencies with possible subsequent loss in high frequency which accompany pathological changes in the cochlea itself

Over the last several decades, there has been an evolution of operative procedures for the correction of deafness resulting from middle ear dysfunction In the earliest operation, Lempert's fenestration procedure, a new window was made from the middle ear cavity to the lateral bony semicircular canal This served to bypass the immobile bone link and partially restore the hearing loss An improvement was achieved by the development of a procedure for mobilization of the stapes by Rosen (Schuknecht, 1962) The method often produces dramatic and immediate improvements in hearing but suffers from the fact that there is a fairly high rate of recurrence of the difficulty The newest procedure is one in which the entire stapes is removed and its footplate replaced by a vein graft or pad of fat over the oval window A linkage to the long process of the incus is accomplished by the use of a wire of plastic prosthesis In cases of hearing deficiency which are the result of chronic *otitis media*, the appropriate procedure may be *tympanoplasty*, the surgical reconstruction of the entire sound-conducting system of the middle ear The precise nature of this operation will vary with the nature of the impairment

A variety of acute effects may result in hearing impairment Several of these cause conduction deafness If the eustachian tube is blocked for any reason during a rapid change in ambient pressure, the balance of pressure between the external meatus and the middle ear will be upset If the middle ear pressure is low and cannot be compensated, as would be the case during descent in aircraft, the drum membrane may be strongly retracted and the conduction efficiency of the system greatly reduced Rupture of the drum will not occur unless the pressure differential reaches the range of 100 to 500 mm of mercury The problem is perhaps more frequently encountered by the diver during descent into deep water A pressure differential which can cause rupture of the drum can easily be reached, but it will generally be preceded by warning pain which will result in the diver's ascent before rupture occurs In spite of the warning which pain provides, transient injury to the drum is a frequent problem for overzealous skin divers

If the eustachian tube remains closed for any length of time, the air trapped in the middle ear is gradually absorbed The lowered partial pressure of oxygen in the cavity may be accompanied by the entrance of fluid into the middle ear when gas pressure falls below hydrostatic pressure in the capillary bed The accumulated fluid increases friction and decreases the transmission of sound

Nerve Deafness. The most common cause of sensorineural impairment is probably exposure to high level sound intensities (Graham, 1966) As mentioned above, there is a temporary shift in threshold (TTS) after exposure to a loud sound or noise (Fig 8 27) If the sound is loud enough and the period of exposure is long enough or repeated frequently enough, the elevation of threshold may be permanent In general, sensorineural deafness affects hearing for higher frequencies to a much greater extent than low frequencies and there appears to be a particular vulnerability in the region of 4000 Hz This may result from the physical nature of the cochlear mechanism which mediates response to sounds of this frequency When caused by exposure to high level sound, such deafness is bilateral When the shift in threshold is of the order of 40 to 60 db , the

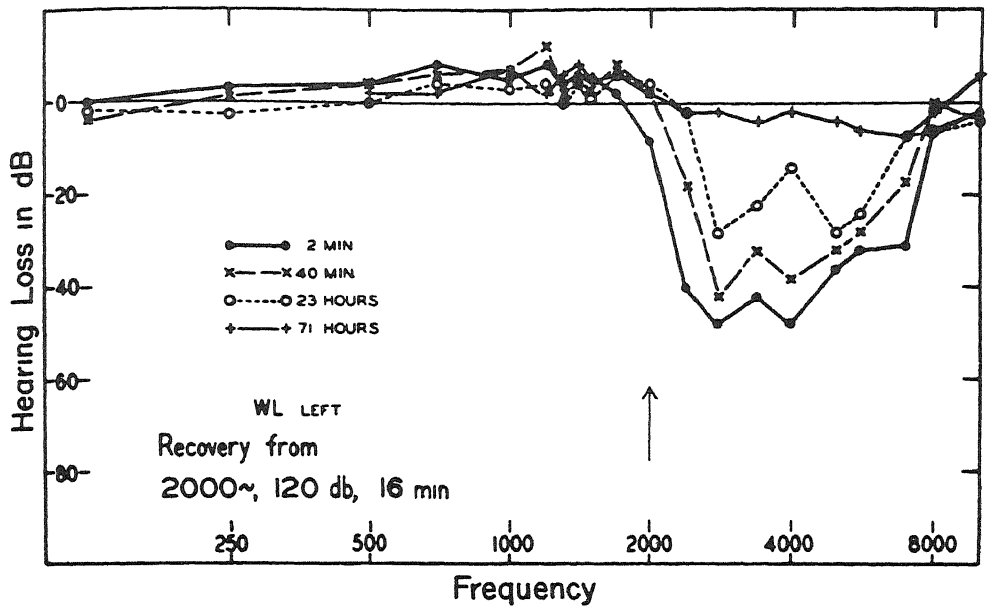


Fig 8 27 Temporary threshold shift (TTS) after exposure to a tone of 2000 Hz at 120 db sound pressure level for 16 min Recovery is shown over the course of time up to 71 hr after exposure Note that the shift of threshold is greatest for frequencies above that of the exposure tone (Davis et al 1950)

loss may be accompanied by *tinnitus*, a sustained ringing sound which is attributed to spontaneous discharge of injured hair cells

Hearing loss associated with injury to the hair cells is accompanied by the phenomenon of *recruitment* That is, at the frequencies of greatest impairment there is a more rapid growth of loudness than normal (compare the rapid growth of loudness in the normal ear at low frequencies) such that the balance of sounds over a broad range of frequencies, including the range of impairment, may be normal at higher sound intensities (Fig 8 28) If sensorineural impairment is the result of infection, it may be unilateral Menière's syndrome is accompanied by hearing loss in the low frequencies, and by sometimes violent attacks of vertigo This probably is the result of an involvement of the other sensory elements in the labyrinth as well as the cochlea Tumors may cause auditory impairment in a variety of ways depending on their location and extent They are not associated with recruitment, however

Finally, just as vision appears to show progressive degradation with increased age, so does hearing (Fig 8.29), particularly for higher frequencies Fortunately, the discrimination of speech depends primarily on frequencies between about 500 and 3000 to 5000 Hz Therefore, the characteristic high frequency loss with increasing age causes only a minimum of difficulty in most circumstances (French and Steinberg, 1947)

A relatively new hazard to hearing results from the specific action of drugs of the streptomycin group on the labyrinthine mechanism They cause degeneration of the hair cells, which may occur even more rapidly for the vestibular apparatus than for the organ of Corti Such drugs have been useful experimentally for pharmacological labyrinthectomy

THE MEASUREMENT AND DIAGNOSIS OF HEARING IMPAIRMENT

Audiometry. The standard clinical tool for the measurement of hearing, the audiometer, consists of an oscillator for the production of selected frequencies usually ranging from 125 to 12,000 Hz , an amplifier, an attenuator or volume control, and headphones which can be driven individually or with various balances of excitation The output of the oscillator is adjusted at each of the frequencies such that when the attenuation control is set at zero, the output of the headphones represents a signal corresponding to the sound pressure level of a

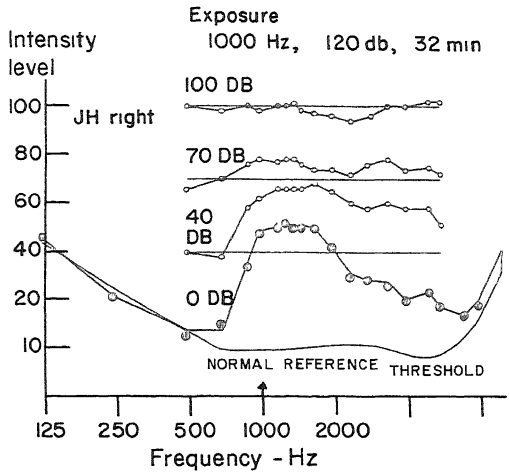


Fig 8 28 Intensity levels required for a match of loudness over the frequency range from approximately 125 to 8000 Hz following exposure to a 1000-Hz tone at 120 db sound pressure level for 32 min Loudness matches are shown for each of four levels relative to the 1000 Hz normal reference level 0 db 40 db 70 db and 100 db Note that there is virtually no elevation of the intensity level required in the frequency range affected for the 100-db level and relatively little effect at the 70-db level (Davis et al 1950)

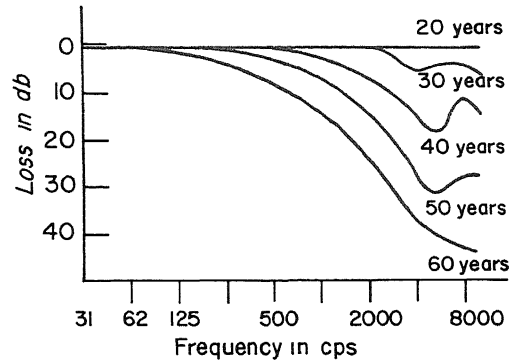


Fig 8 29 Progressive loss of sensitivity at high frequencies with increasing age The normal audiogram at 20 years of age serves as a reference for the losses represented at greater ages (Licklider 1951)

standard reference curve (see Fig 8 24) The attenuation control is usually calibrated in 5-db steps An increase in the reading of the attenuation control corresponds to an increase in the sound power level at the headphones by an equivalent amount Thus, if an appropriate attenuator setting can be determined at which a subject just hears the tone at a given frequency, then the reading of the attenuator control represents the sound power level over a normal standard which is required for that patient at that frequency An audiogram can then be plotted which represents hearing loss at each of the selected frequencies in terms of the increased sound power required in decibels as compared with the normal level The examiner need not concern himself with the specific form of the frequency response of the ear The device is compensated for this and provides for a direct measurement of loss relative to the appropriate reference at each frequency tested

There are a variety of procedures for testing hearing with an audiometer (Jerger, 1963) A fairly simple one for most patients is the presentation of tones continuously at a level slightly above threshold The continuous tone is interrupted at random intervals by the examiner The patient is instructed to signal the interruption of the tone by depression of a switch which causes a light to come on on the instrument panel The whole procedure has been automated in a device designed by von Bekésy in which the patient holds a button depressed as long as he hears the tone and releases it when he is unable to hear the tone Depression of the button drives the output of the oscillator down Output increases at a regular rate when the button is not depressed The frequency spectrum is swept slowly from 100 to 10,000 Hz and a recording pen traces the upward and downward changes in intensity corresponding to the subject's manipulation of the switch The midpoint between reversals of direction of the pen may be taken as a reasonable representation of threshold Examples of audiograms obtained with the conventional procedure are shown in Figure 8 30, a von Békésy audiogram is illustrated in Figure 8 31

Several methods have been developed for the testing of hearing without the requirement for any instruction or special performance on the part of the patient One of these depends upon an *electrodermal response* (EDR) or change in resistance of the skin which follows a mild electric shock The shock is regularly paired with the testing tone until it comes to evoke the EDR, even before shock is presented In another test, the electroencephalographic response recorded from scalp electrodes provides an indication of whether a tone has been heard Responses are evoked in the electroencephalogram by sound stimulation for

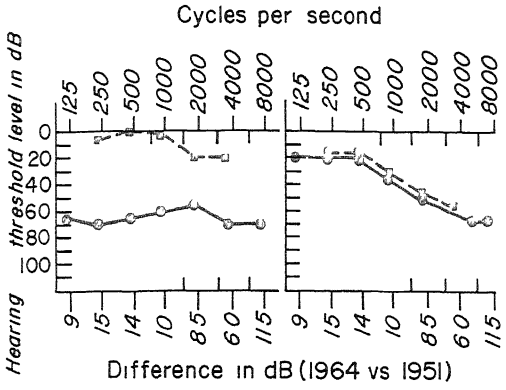


Fig 8 30 Audiograms illustrating conductive impairment of hearing on the left and sensorineural impairment of hearing on the right The solid curve is for air conduction the dashed curve for bone conduction Note that bone conduction and air conduction curves are quite similar in sensorineural impairment

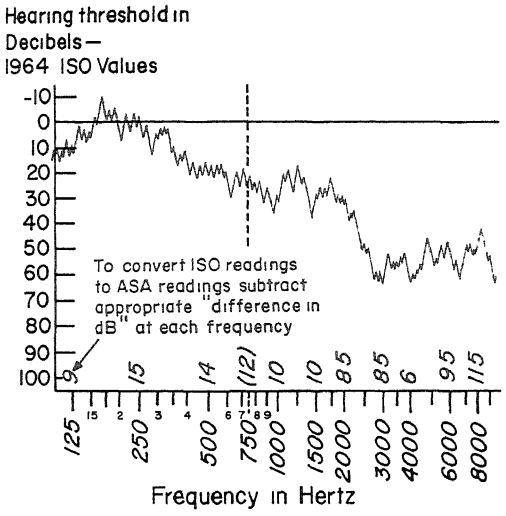


Fig 8 31 A von Bekesy audiogram of a middle-aged patient with some sensorineural impairment attributable to cumulative effects of quinine ototoxicity repeated experimental exposures to sound at 120 to 140 db sound pressure level and advancing age

appropriately placed electrodes, and a skilled operator may determine threshold values by the interpretation of such records

Bone Conduction Tests. A variety of procedures is available which assist in the diagnosis of hearing impairment In addition to the headphones used with an audiometer, a small vibrator may also be used This is placed either on the forehead or on the mastoid prominence behind the ear Vibrations of the skull are transmitted to the cochlea and oval window directly For frequencies up to about 1500 Hz , the nature of suspension of the ossicles in the middle ear is such that they do not tend to move with skull vibration Thus, there is a relative motion between the footplate of the stapes and the oval window Sound is heard just as it would be if the stapes itself were vibrating At higher frequencies, the skull does not vibrate as a rigid body but rather in sections As vibrations are transmitted to the bony walls of the cochlea, the contained fluids are set in motion and hair cells are stimulated With either mechanism of bone conduction, the pathways of the external meatus and the middle ear are bypassed Thus, a loss which may be registered when measurements are made with air conduction by reason of conduction deafness will not appear Of course, account must be taken of the lower efficiency of the bone conduction pathways to the inner ear The difference in threshold may be anywhere between 30 and 60 db , depending on the conditions of measurement A special attenuation scale is often included on a clinical audiometer which differs from the scale employed with headphones by approximately 30 db , at least over a limited range of calibrated frequencies The audiograms illustrated in Figure 8 30 provide an example of the diagnostic value of employing both air conduction and bone conduction measurements In the curve on the left, a near normal audiogram is obtained by bone conduction while there is a better than 60 db attenuation with air conduction Clearly, the nature of the impairment here involves conduction to the inner ear On the other hand, in the figure on the right, the roughly equal impairments demonstrated by air conduction and bone conduction signify some form of sensorineural involvement Problems may arise in testing unilateral deafness with bone conduction techniques Sound may be conducted around the head to the good ear when the bone conduction vibrator is located on the mastoid bone behind the impaired ear The patient can then provide a correct report as to the presence or absence of the tone near normal levels in spite of his impairment This problem is overcome by presenting a masking sound to the good ear As indicated above, the contralateral effect of masking is relatively minimal The appropriate masking level may be adjusted on the basis of comparative meas-

urements of air conduction thresholds in each of the ears When the good ear is masked, bone conduction thresholds can confidently be made on the impaired ear

Diagnosis of Recruitment Recruitment, or an abnormally rapid increase in loudness with increase in sound intensity, can be demonstrated in cases of unilateral hearing impairment by permitting a patient to adjust the balance control of the audiometer so that the sound heard appears to be in a medial plane or centered in the head When the loudness is equal for the two ears, this will be the case At very low levels, it will be necessary to increase the excitation in the phone over the impaired ear with respect to the level of excitation in the phone over the good ear As the sound level is raised, the amount of imbalance decreases until at somewhere in the neighborhood of 70 db the imbalance may be completely eliminated

Other Tests of Hearing Impairment There is a variety of relatively simple diagnostic tests of hearing impairment which can be made with nothing more complicated than a tuning fork At the other extreme, techniques for the measurement of the impedance presented by the external canal of the ear with its termination at the tympanic membrane have been developed These permit the identification of changes which result from pathology involving the drum or the ossicular chain Application of a loud sound to the opposite ear will elicit a reflex contraction of the stapedius muscle of the ear under test which, in the normal ear, will result in a characteristic change in the measured impedance Recruitment may be observed if a normal stapedius reflex is obtained at a relatively high stimulus level in spite of an elevated threshold

Speech Audiometry For practical purposes, particularly the extent to which understanding of speech may be affected by a hearing loss, the absolute threshold for the entire range of frequencies may be of little concern and even the absolute threshold in that range of frequencies which is most important for the discrimination of speech may be of little importance As mentioned above, in certain types of sensorineural deafness, there is a recovery of loudness up to near normal level with increase in the intensity of a sound above threshold, even at a frequency where there may be gross impairment based on threshold determinations For this reason, a variety of tests of hearing based on the recognition of spoken words has been developed Selected test words may include spondees, words of two

Table 8 2
Classes of hearing handicap ISO–1964

db	Class	Degree of Handicap	Average Hearing Threshold Level for 500, 1000, and 2000 Hz in the Better Ear		Ability to Understand Speech
			More than	No more than	
	A	Not significant		25 db (ISO)	No significant difficulty with faint speech
25					
	B	Slight handicap	25 db (ISO)	40 db	Difficulty only with faint speech
40					
	C	Mild handicap	30 db	55 db	Frequent difficulty with normal speech
55					
	D	Marked handicap	55 db	70 db	Frequent difficulty with loud speech
70					
	E	Severe handicap	70 db	90 db	Can understand only shouted or amplified speech
90					
	F	Extreme handicap	90 db		Usually cannot understand even amplified speech

From Davis (1965b)

syllables which are equally stressed, or monosyllabic words Lists of words are presented to the patient from phonographic or tape recordings such that the loudness can be standardized and controlled Performance is based on the number of words that can be repeated correctly at various intensity levels Lists of words are selected to include appropriate representation of the various speech sounds in conversational English The percentage of the words which the patient is able to repeat correctly is his discrimination score or articulation score An index of the adequacy of hearing for normal social interaction may be derived from such scores (Davis, 1965)

In the absence of elaborate tests of hearing which involve the spoken word, the elevation of threshold at frequencies of 500, 1000, and 2000 Hz provides a useful indication of the significance of the impairment (See Table 8 2)

HEARING AIDS

Hearing aids improve the conduction of sound to the inner ear They may thus be useful in aiding someone with a conduction deafness or even with sensorineural impairment when the resulting deafness is not complete No effort is made to match the response characteristics of the hearing aid to the frequency response characteristics of the loss of the patient Such "prescription" correction has not been found to be of any benefit With severe sensorineural defects, there may be little benefit derivable from the use of any type of hearing aid

3 / VESTIBULAR FUNCTION

The vestibular mechanism is an extremely sensitive sensory device that provides information as to the nature of movements of an organism and its position in space The sensory receptors of the vestibular system are activated by acceleration forces The semicircular canals are uniquely suited to the detection of angular acceleration components in three dimensional space while the utricles and the saccules are so constructed that they will respond to linear accelerations The utricles are believed to play the predominant role in the determination of the posture of the head under static conditions The stimulus for this kind of discrimination depends upon the orientation of the earth's gravitational acceleration vector with respect to the sensory organ The saccules are richly innervated and in many ways quite similar to the utricles, but their precise function is somewhat uncertain It has been suggested that they may be of most importance for the detection of vibration and possibly very low frequency sound

The vestibular sense does not have the conscious prominence of vision, audition, touch, taste or smell Normally, one is unaware of any vestibular dimension of sensory experience Only with malfunction of the system with resulting dizziness, nausea and other symptoms does the function of the vestibular mechanism attain prominence

Vestibular signals are undoubtedly of great importance for an animal as it moves about in space The control of movements of the entire body, of the head, and particularly of the eyes with respect to the head is refined by feedback signals from the vestibular mechanism in response to bodily movement There are clearly very important species differences with respect to the prominence of the vestibular mechanism Animals which live in three dimensions, such as birds and fish and climbing animals including the cat, have highly developed neural support systems for the vestibular detectors and these animals tend to be dependent to a greater degree than other species on the normal function of this system

We tend to think of the world as we perceive it, at least when we are studying sensory processes, as a world made up of dimensions that correspond to our various sensory systems For man, the vestibular sense hardly has any independent identity by reason of its lack of any conscious concomitant most of the time In fact, our perceptions of the world depend in large degree on interactions among the senses, and the stability of the visual world in particular is dependent upon the coordination of motor responses involving postural mechanisms, the extraocular muscles and sensory information provided by the vestibular mechanism

Anatomy and Physiology of Labyrinthine Systems

COMPONENTS OF THESE SYSTEMS

The extra-auditory portion of the labyrinthine system includes the three semicircular canals, the utricle and the saccule (Fig 8 32) The sensory endings in these structures, located in the cristae of the canals and the maculas of the utricle and saccule, are all innervated by fibers of the vestibular division of the eighth cranial nerve These organ systems are made up of interconnected fluid-filled ducts and sacs which are enclosed in the bony labyrinth within the temporal bone The semicircular canals open at each end to the vestibular chamber in which are located the utricle and saccule This portion of the system is continuous with the auditory labyrinth The entire membranous labyrinth is filled with *endolymph* similar to intracellular fluid and of relatively high potassium concentration Surrounding the membranous system within the bony labyrinthine cavity is the perilymph, similar in composition to the cerebral spinal fluid The nonauditory labyrinth is distinguished from the entire labryinthine system

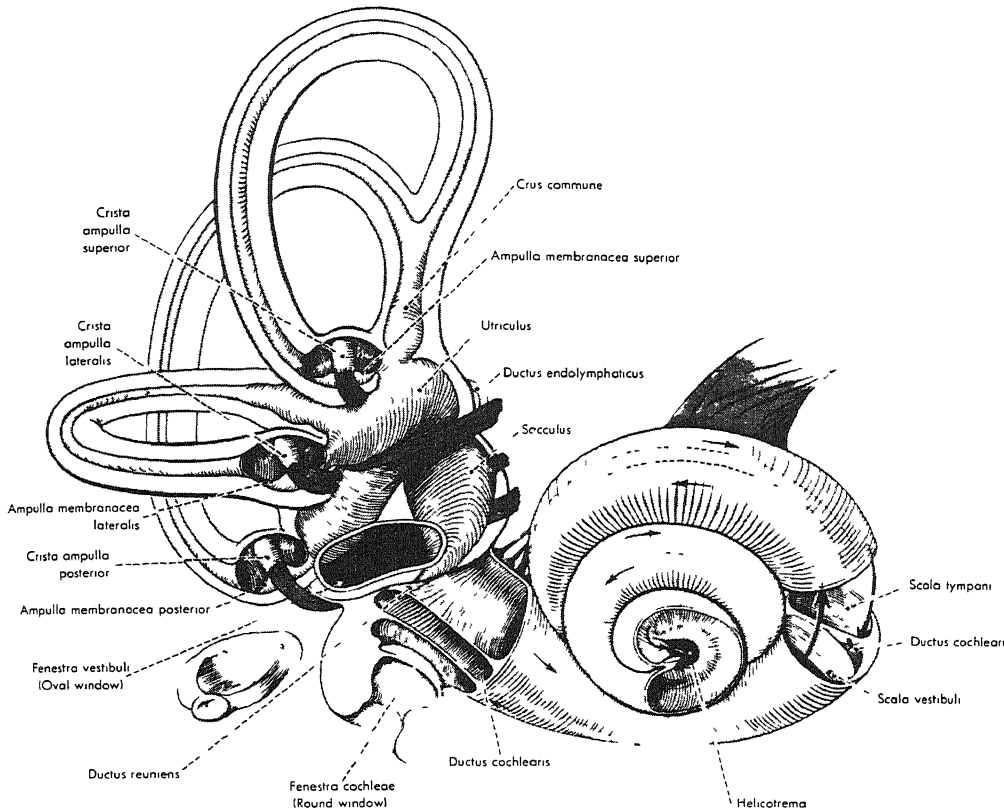


Fig 8.32 A schematic illustration of relations among the semicircular canals the utricle and the saccule and the cochlear duct of the right side The membranous semicircular canals are shown within their surrounding bony structure Nerve connections into the ampullae and the utricle and saccule are shown Various parts of the anatomy of the entire mechanism are identified on the figure (Melloni 1957)

including the cochlea as the *vestibular system* There are several excellent detailed treatments of the anatomy and physiology of this system which provide more information than the present account (Wendt, 1951, Howard and Templeton, 1966, Roberts, 1967)

The Semicircular Canals. The three semicircular canals on each side of the head are oriented in each of three planes which are very nearly perpendicular to each other (Fig 8 32) This arrangement permits response to angular accelerations about any axis in three dimensional space The sensory epithelium of the semicircular canals is located within the ampulla of each canal (Fig 8 33) It is evident (Fig 8.32) that the ampulla of the horizontal or lateral canal, and that of the superior or anterior canal are located close together near the anterior terminations of these canals on the vestibule in the region occupied by the utricle The ampulla of the posterior or inferior canal is proximal to the opposite side of the utricle near the narrow end of the horizontal canal The two vertical canals, superior and posterior, join together in the common crus prior to their entry into the utricular portion of the vestibule

When the head is held erect the horizontal canals are actually inclined downward toward the rear by approximately 30° If the head is tilted forward such that the opening of the external meatus and the outer canthus of the eyelids are on the same horizontal line, then the horizontal canals are very nearly in a horizontal plane (Fig 8 34) Since the anterior canal on one side is in a plane parallel to that of the posterior canal on the other side, the contralateral anterior and posterior canals are maximally stimulated by rotations about a common axis

Stimulation of a given pair of semicircular canals (the horizontal canals in the same plane

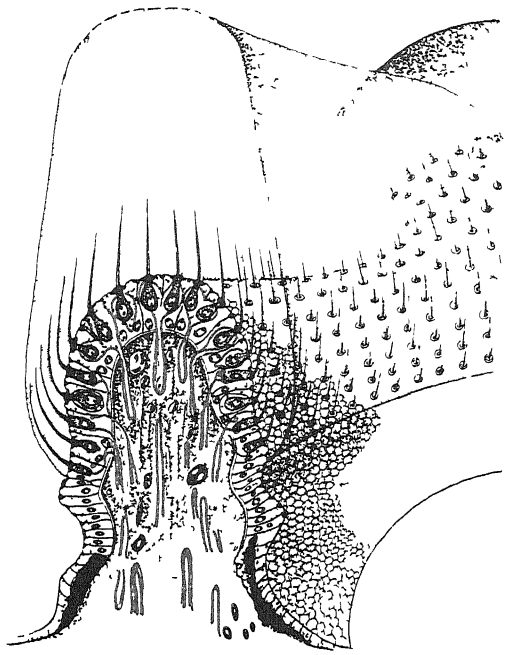


Fig 8 33 Illustration in three dimensions of the crista ampullaris (Wersäll 1956)

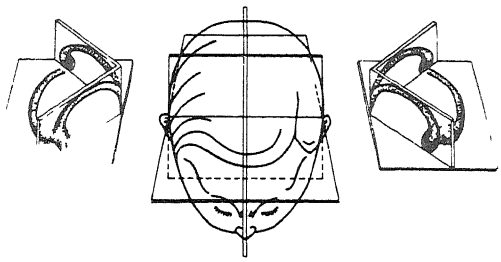


Fig 8 34 Relations of the semicircular canals to the median sagittal transverse frontal and horizontal planes of the head

or vertical canals in parallel planes) occurs whenever a component of angular motion is introduced about an axis which is perpendicular to their plane or planes The mechanism of excitation is bending of the cupula This is caused by the inertial displacement of the endolymph within the membranous canal whenever there is any angular acceleration of the canal The fluid tends to remain stationary as the motion of the canal commences The cupula is thus bent about its point of attachment to the wall of the ampulla as the wall moves at a greater initial rate than does the fluid which the canal contains The cupula, a gelatinous mass extending from the top of the crista to the opposite wall of the ampulla, is in sufficiently close contact with the wall to block the flow of any fluid around it This has been demonstrated by the injection of dye into the canal in fish and the observation that with rotation of the system and bending of the crista and cupula there is no leakage of dye from one side to the other (Steinhausen, 1933) Bending of the cupula causes bending of the hair cells which arise out of the crista and are enveloped in the gelatinous substance of the cupula

If the organism is rotated at a constant velocity, the endolymph within the rotating semicircular canals reaches the same velocity as the rest of the system, although belatedly, and excitation of the receptor endings ceases If the rate of rotation is then reduced, the inertia of the endolymph tends to keep it in motion as the canal slows down and there is bending of the cupula in the opposite direction

Apparently, the mucopolysaccharides which comprise the cupula are secreted by the supporting cells of the crista There are two types of hair cells on the crista (Fig 8 35) The cylindrical type is served by both afferent and efferent nerve endings

Each of the hair cells has from 60 to 100 stereocilia or individual hairs which emerge from its distal surface A single *kinocilium* is located on one side of the cluster of stereocilia The kinocilium is the longest of the cilia and the stereocilia are graded in length with the longest being closest to the kinocilium The two types of hair cells are distributed over the entire surface of the crista, but the bottle-shaped cell (Fig 8 35) is concentrated on the vertex of the crista In the cristae of the horizontal canals, the kinocilia of the hair cells are always located on the side toward the utricle On the other hand, in the case of both the anterior and the posterior vertical canals, the kinocilia are always located on the side of the hair cells away from the utricle It has been demonstrated that bending of the cilia of the hair cells in the direction of the kinocilium results in depolarization of the cells and heightened neural activity Bending in the opposite direction results in hyperpolarization and reduction of the base level of activity (Fig 8 36) Thus, when patterns of motion are such that the endolymph tends to move with respect to the membranous wall of the canals, there will be excitation in the member of a pair of canals on one side of the head and inhi-

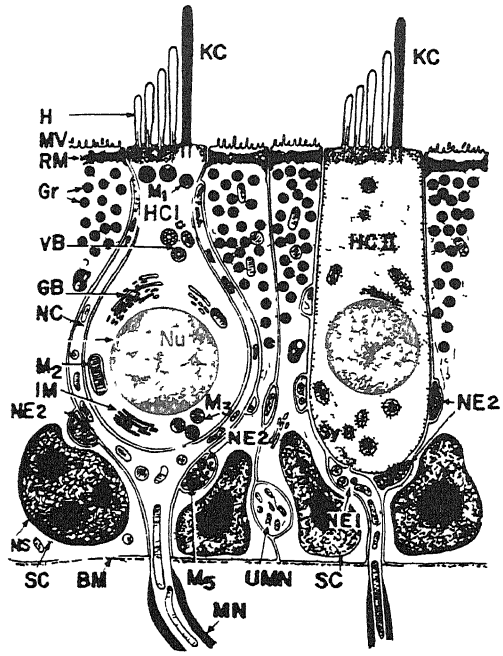


Fig 8 35 Two types of vestibular sensory cells The flask-shaped type I cell (HC I) is surrounded by a nerve calyx (NC) Nerve endings presumed to be efferent (NE 2) make contact with the nerve calyx and directly with the type II cylindrical sense cell (HC II) These cells are supplied by both NE 1 and NE 2 types of nerve endings Several types of mitochondria (M_1 to M_5) are found in the sensory cells and neural elements Additional elements identified in the figure include kinocilia (KC) stereocilia or hairs (H), microvilli (MV) granules (Gr) cell nuclei (Nu) supporting cells (SC) unmyelinated nerve fibers (UMN) myelinated fibers (MN) basement membrane (BM) Golgi complex (GB) synaptic bar (SyB) reticular membranes (RM) intracellular membrane (IM) vesiculated body (VB) and nucleus of supporting cell (NS) (Engström Ades and Hawkins, 1965b)

bition on the other side Rotation of the body to the right about a vertical axis with head erect will cause bending of the cupula in the right horizontal canal toward the utricle or vestibule This will result in excitation The horizontal canal on the left side will be inhibited as its cupula is bent away from the opening into the utricular chamber

Depending upon the orientation of the axis of angular acceleration, the two anterior vertical canals may both be inhibited, both excited, one inhibited and one excited or neither stimulated It must be remembered that they do not operate as a pair, however The anterior canal on one side and the posterior canal on the other work as a pair and they are so arranged that whenever one is excited, the other will be inhibited In the case of both the anterior and posterior vertical canals, rotations which result in bending of the cupula away from the opening into the utricle cause excitation

The lateral selectivity to direction of angular rotation was first established by Ewald (1892), who stated the nature of the response in what is known as Ewald's Second Law The subsequent discovery of the kinocilia and their relative orientations in the different canals provides us with a knowledge of the anatomy which underlies this law (Ewald's First Law stated that the cupula within the ampulla was the site of sensory activation)

The diagonal orientation of the two vertical canals results in both anterior and posterior canal stimulation for rotation about either a transverse lateral or a transverse medial axis Only rotations about an axis perpendicular to the parallel planes of the anterior canal on one side and the posterior canal on the other side

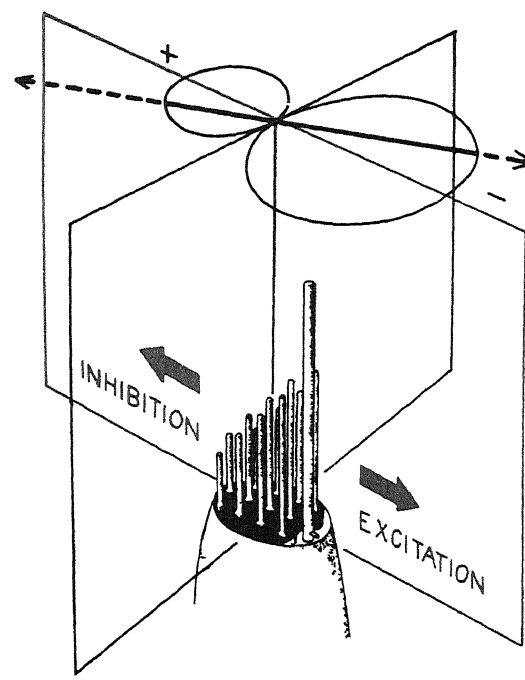


Fig 8 36 The excitation depolarization response which results when hair cells are bent in the direction of the kinocilium or the inhibition polarization response for bending in the opposite direction For a given direction of bending the amplitude of effect is approximated by the length of a line drawn from the point of tangency of the two circles to the intersection of either the circle representing excitation or the circle representing inhibition (Flock 1965)

would fail to stimulate more than one of the vertical canals The backward tilted position of all three of the canals when the head is erect results in excitation of two vertical canals and one horizontal canal for any rotation about a vertical axis The pattern of excitation and inhibition of the six canals is shown in Table 8 3 for rotations about longitudinal, transverse and vertical axes

Excitation of the appropriate canals commences with the beginning of angular acceleration and continues for twenty or more seconds after a constant angular velocity has been reached Neural discharges signaling excitation begin again upon decrease in angular velocity, this time from the opposite canals of the various pairs stimulated by the original rotation Various factors which influence the threshold angular acceleration for neural activity are discussed below Suffice it to say here that the threshold angular rate of acceleration is of the order of ½° per sec ^{2} (Clark, 1967)

The Otolith Organs. The function of otolith organs is fundamentally different from the function of semicircular canals Whereas the latter are excited by reason of the inertial properties of endolymph within the membranous canals, in the otolith organs, stimulation depends upon the difference in density of the otoconia or statoconia, small calcium carbonate crystals, with respect to the gelatinous substrate which supports them The hair cells of the otolith organ are virtually identical with those of the cristae of the semicircular canals Deflection of the hair cells of the otoliths depends upon relative motion of the otoconia with respect to the supporting substrate The cilia of the otolith hair cells are somewhat shorter than those of the semicircular canals (5 to 7 μ) and extend into the gelatinous substrate (Fig 8 37) The specific gravity of the otoconia is approximately 2 94 This is considerably higher than the specific gravity of the supporting structure Thus, the action of a component of linear acceleration on the otoconia results in the exertion of a force for deformation of the otolithic membrane and consequently, possible bending of the hair cells The direction of bending will depend upon the orientation of the resultant acceleration force Terrestrial creatures are continuously subject to the acceleration of the earth's gravitational field and any movement of the head may change the orientation of this acceleration with respect to the otolith organs Thus, the direction of bending of hair cells will vary with head position

Table 8 3
Responses of the six semicircular canals (of fish the thornback ray) to rotation

Semicircular Canal	Rotation about the: Longitudinal axis — Right	Longitudinal axis — Left	Transverse axis — Forwards	Transverse axis — Backwards	Vertical axis — Clockwise	Vertical axis — Anticlockwise
Right anterior vertical	●	⊗	●	⊗	⊗	●
Left anterior vertical	⊗	●	●	⊗	●	⊗
Right posterior vertical	●	⊗	⊗	●	●	⊗
Left posterior vertical	⊗	●	⊗	●	⊗	●
Right horizontal	○	○	○	○	●	⊗
Left horizontal	○	○	○	○	⊗	●

● excited, ⊗ inhibited, ○ unaffected (After Lowenstein and Sand, 1940)

The otoliths have been called static receptors in contrast to the semicircular canals which are called dynamic receptors For a given position of the head, otoconia may exert a force which will maintain hair cells in a bent position, even though the head is maintained stationary In the case of the semicircular canals, a force acts upon the cupula only during angular acceleration At constant velocity of rotation, the cupula returns to its resting position and excitation of the hair cells ceases At least some cells of the otolith organs respond continuously for a given head position There would appear to be a characteristic pattern of nerve fiber activity associated with any specific position of the head

The Utricles. The utricular organ (Fig 8 32) communicates with the saccule by way of the utriculosaccular duct As mentioned above, the circular loop of the semicircular canals is closed through the utricle The macula of the utricle is similar to the crista, with both types of hair cells and supporting cells The cupula is replaced by the otolithic membrane, however, into which extend the shorter hairs of the utricular hair cells The densely packed utricular stones or calcite crystals are embedded in the upper surface of the otolithic membrane The macular surface of the utricle is



Fig 8 37 Bundles of sensory hair cells at the surface of the utricular macula as revealed by the scanning electron-microscope A freeze-drying technique was employed for preparation of the specimen (Engström, 1968)

located on the lower and forward walls of the utricular sac with its posterior two-thirds in a plane approximately parallel to that of the horizontal canal The anterior third is turned up slightly (Fig 8 38) The right and left utricular maculas are in the same orientation and are symmetrical in shape and organization

Unlike the hair cells of the cristae, the hair cells of the utricular maculas do not all have the same orientation with respect to location of their kinocilia Virtually every orientation is represented in each of the macular surfaces The arrows in Figure 8 39 represent orientations of the hair cells over the surface of the

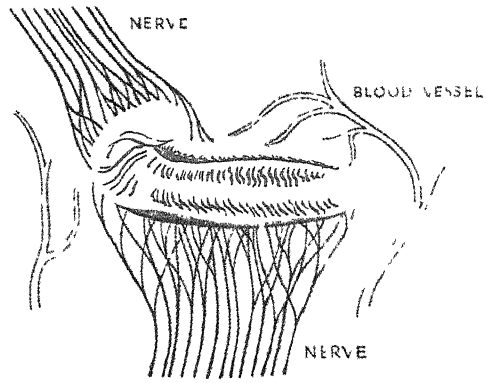


Fig 8 38 The utricular macula of the guinea pig showing separate innervation of its two portions one from the upper and one from the lower division of the vestibular nerve There is presumably also some efferent innervation (Ades and Engstrom 1965)

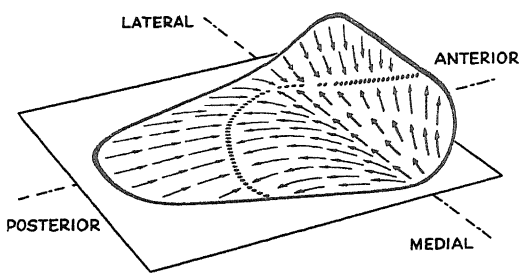


Fig 8 39 The pattern of polarization of sensory cells in the utricular macula of the guinea pig Arrows indicate the direction of bending of statocilia for excitation (Spoendlin 1965)

macula It is evident that tilting of the head from a vertical position in any direction will result in the bending of hairs of at least some of the hair cells in the direction of their kinocilia There would thus appear to be possible a kind of position coding of excitation on the surface of the macula corresponding to the position of the head If the head is tilted backward and down through 150°, most of the otoconia will be pulled out and away from the gelatinous substrate of the macula by the force of gravity As one might expect, under these circumstances, the ability of a human subject to correctly identify his head position in the absence of other cues is least accurate (Brown, 1960) Electrophysiological studies (Lowenstein and Roberts, 1951) have revealed that tilting of the head about any horizontal axis gives rise to activation of the macula of the utricle on either side of the body Individual functional units give maximal responses in accordance with the kind of orientation pattern depicted in Figure 8 39 In addition to the response of the system to the orientation of the linear acceleration vector of the earth's gravitational field, there is a displacement of the otoconia whenever the body is subjected to linear accelerations of any sort under dynamic conditions In addition, rotation about any axis which does not pass through the utricle is accompanied by a linear acceleration component, the instantaneous turning moment of which displaces otoconia and bends hair cells

The dotted line in Figure 8 39, which illustrates the boundary between units of reversed polarity, is also associated anatomically with a boundary between relatively large otoliths on one side and smaller ones on the other (Spoendlin, 1965)

The Saccules. The saccule is connected to the cochlear duct by way of the ductus reuniens The macula of the saccule is located on the medial wall of this cavity in an approximately vertical orientation in an anteroposterior plane The otoliths (Fig 8 40) are lateral, with the macular substrate medial There is a slight concavity of the macular surface as viewed from its lateral aspect on a given side The surface and its shape along with the direction of orientation of the kinocilia of the hair cells are illustrated in Figure 8 41. Unlike the macular surface of the utricle, the hair cells on either side of a line representing a reversal in orientation have kinocilia oriented away from each other rather than toward each other In spite of the striking similarity between the saccules and the utricles, the nature of the function of the saccules is still not understood The nature of the structure of the saccules suggests that they are probably responsive to linear accelerations, but electrophysiological confirmation of any understandable functional role for these organs remains to be accomplished.

NEURAL INNERVATION OF VESTIBULAR EPITHELIA

The diameter of fibers ending on the type 1 hair cells that have the shape of a flask or a chalice (Fig. 8 36) is thought to be, in general, greater than the diameter of fibers terminating

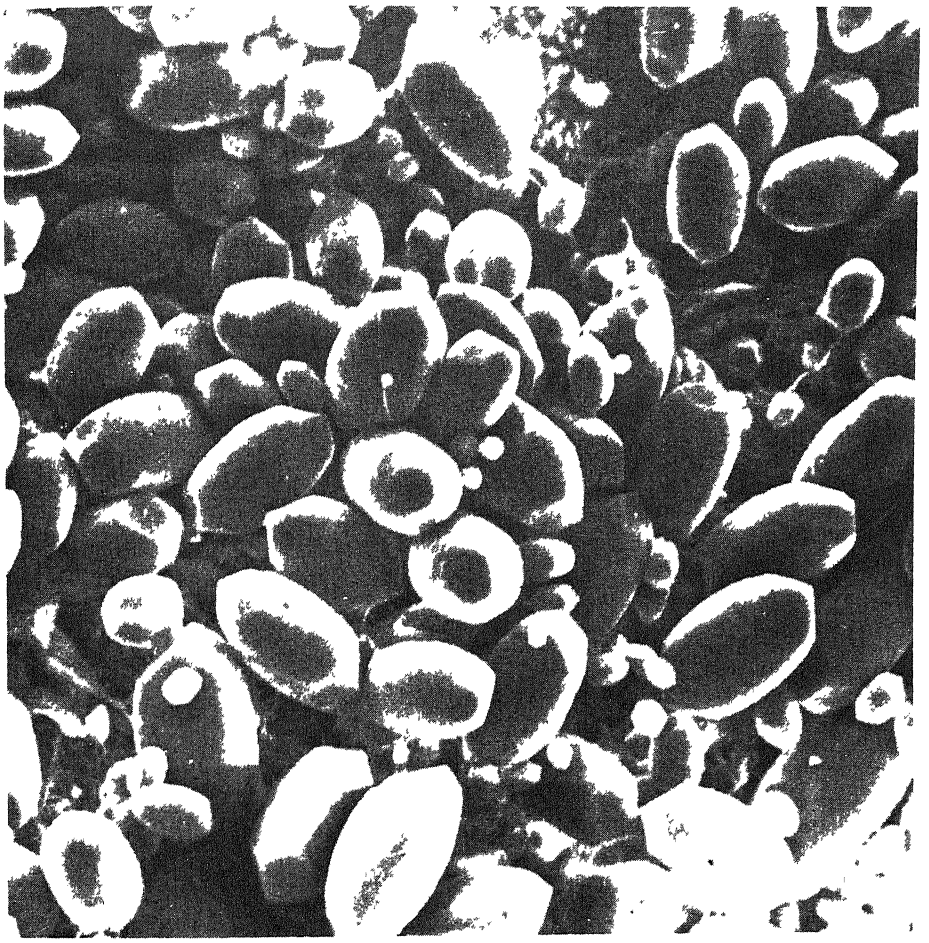


Fig 8 40 Otoconia of the macula of man as revealed by a scanning electron microscope Hexagonal prisms may be discerned clearly in unsectioned material (Photomicrograph provided through the courtesy of Hans Engstrom See Engstrom 1968)

on the cylindrical or type 2 hair cells (Wersall, 1956, Engstrom and Wersall, 1958) It has also been shown, however, that cells of both types may be innervated by branches of the same nerve fiber Innervation of the type 2 cell may come directly from a nerve calyx surrounding a type 1 cell The type 2 cells receive at least two different types of nerve endings The first of these forms a synaptic contact with the hair cell plasma membrane which may be somewhat indented into the cell itself (Engstrom, 1961) These endings are believed to be those of afferent fibers

Another type of nerve ending on these cylindrical hair cells is bouton-shaped, densely granulated and in direct contact with the cell surface There is evidence that nerve endings of this type also contact nerve fibers of the former type at synaptic junctions The chalice-shaped hair cell is almost completely enveloped by a sparsely granulated nerve calix Recent investigations have shown that there are areas where the space between the calix and the cell is reduced which may be associated with a synaptic bar (Ades and Engstrom, 1965) Densely granulated endings of the type associated with the cylindrical cells are found outside the nerve calix (Engstrom, 1961) Collateral branches between the cylindrical and chalice-shaped cells are common Sometimes, several hair cells may be found enclosed in a single chalice (Ades and Engstrôm, 1965)

Studies in which the vestibular nerve or the vestibular nuclei have been transected indicate that the vesiculated nerve endings are efferent

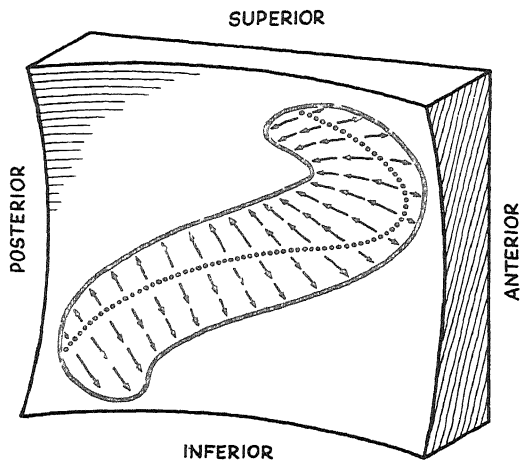


Fig 8 41 The polarization pattern of sensory cells in the saccular macula of the guinea pig Note that although the direction for depolarization is reversed in the region of the middle of the structure as was true for the utricle the direction of bending for excitation is reversed as contrasted with that shown in Figure 10 (Spoendlin 1965)

and that the nerve fibers to which they correspond enter with the vestibular nerve (Smith and Rasmussen, 1963) Selective Wallerian degeneration is most notable in six to ten days after section of the nerve root These fibers show selective staining for acetylcholinesterase which quality is lost after sectioning The fibers have therefore been compared with the efferent fibers to the cochlea in the olivocochlear bundle The efferent fibers apparently form a horizontal nerve plexus which tends to be found in the lower half of the sensory structures (Fig 8 42)

Although there is a difference in density of the neural innervation, the density being greater for the maculas, the arrangements are similar for maculas and cristae At least some of the neurons which give rise to efferent processes appear to be located in the lateral vestibular nucleus (Gacek, 1961) The number of vestibular efferent fibers is estimated to be ap-

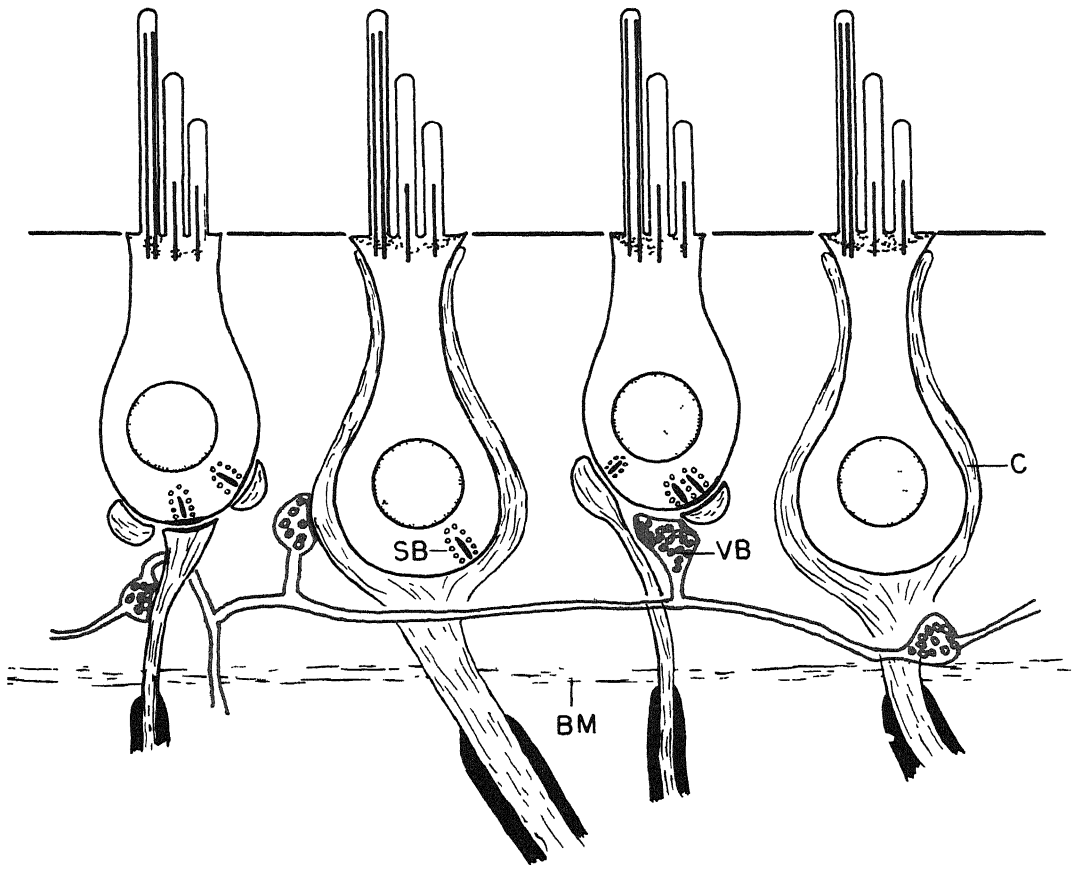


Fig 8 42 Four hair cells and their nerve endings shown in relation to vesiculated boutons (VB) chalice terminals (C) and other boutons and nerve fibers in the chinchilla macula BM basement membrane SB synaptic bar The horizontal plexus is believed formed by efferent nerves (Smith and Rasmussen 1967)

proximately 200 Their pathways are illustrated in Figure 8 43 It is fairly well established that they permit direct inhibitory effects on sensory elements of the receptor mechanism (Sala, 1968, Wersall, 1968)

CENTRAL NEURAL CONNECTIONS OF VESTIBULAR RECEPTORS

Afferent fibers of vestibular receptors are conveyed to the medulla by the vestibular branch of the eighth nerve The cell bodies of the bipolar vestibular neurons are in Scarpa's ganglion within the internal auditory meatus Axons terminate in the superior, lateral, medial and inferior vestibular nuclei of the medulla as well as in the flocculonodular lobe and the fastigial nuclei of the cerebellum Fibers from the superior nucleus join the medial longitudinal bundle on the same side Those from the medial nucleus cross to the medial longitudinal fasciculus on the opposite side Fibers from the inferior nucleus ascend in the medial longitudinal fasciculus on both sides Fibers from these nuclei terminate in nuclei of the third, fourth and sixth cranial nerves They constitute the vestibulo-ocular tract which mediates reflex movements of the eyes (Fig 8 44)

Descending fibers from the inferior nucleus comprise a portion of the spinal extension of the medial longitudinal fasciculus The fibers of the lateral nucleus comprise the descending vestibulospinal tract Postural reflexes involving the neck, trunk and limbs are mediated by these pathways as illustrated in Figure 8 45*A* Connections to the vestibulocerebellar tract are illustrated in Figure 8 45*B* There are a variety of connections between vestibular nuclei and reticular formation which are probably involved in the modulation of flexion and extension responses of the limbs as well as in the activities of other sensory systems

METHODS OF STUDY OF THE VESTIBULAR SYSTEM

A variety of techniques has been employed by physiologists in their efforts to investigate the vestibular system For stimulation of the semicircular canals, rotation is the most appropriate and the most frequently employed procedure (as in the Bárány chair which permits rotation about a vertical axis) Another procedure useful clinically is caloric stimulation Water of a higher or lower temperature

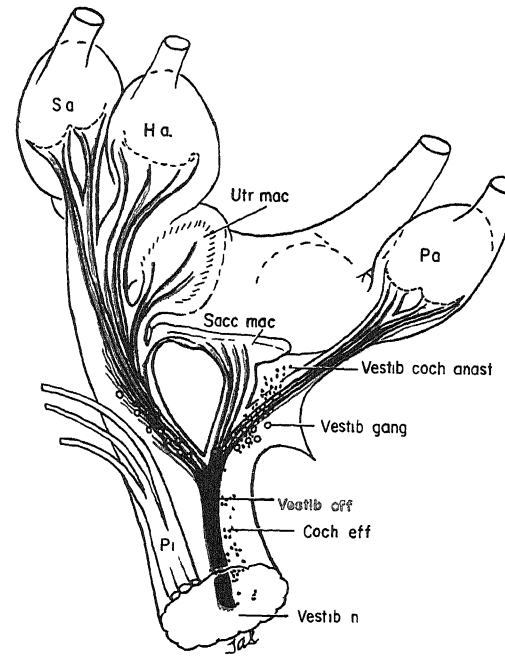


Fig 8 43 Diagram of the vestibular nerve and branches illustrating the course and distribution of the efferent vestibular fibers to the superior ampulla (S a) the horizontal ampulla (H a) and the posterior ampulla (P a) (Gacek 1967)

than body temperature may be injected into the external meatus for varying intervals of time If the temperature gradient is sufficient and a long enough duration of stimulation is employed, the temperature of the fluids of the semicircular canals is apparently raised or lowered in the region of the vestibule resulting in convection currents which serve to stimulate the canals A high temperature solution may be used on one side, and a low temperature on the other to maximize effects On the other hand, one value of the technique is that it permits selective stimulation of the semicircular canal system on one side of the body The supposition that convection currents are in fact responsible for the effectiveness of caloric stimulation is supported by a study which indicates that caloric stimulation is uneffective under conditions of zero gravity (Kellogg and Graybiel, 1967) Without a gravitational force, there are no convection currents

Stimulus. Stimulation of the otolith systems can be achieved by the simple expedient of repositioning the head in the earth's gravitational field Responses associated with otolith organs have also been studied under reduced

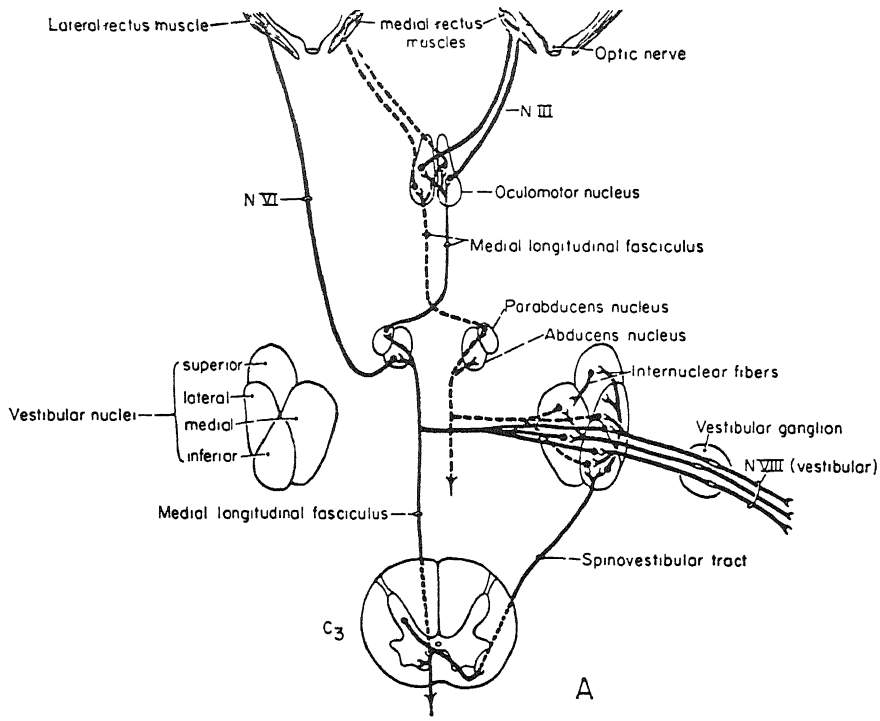


Fig 8 44 Relation of the vestibular nuclei to the medial longitudinal fasciculus and the motor nuclei of the eye muscles for movements in the horizontal plane (Crosby Humphrey and Lauer 1962)

gravitational acceleration in ballistic trajectory flights which can afford brief intervals of zero gravity Effects of increased gravitational force may be investigated on linear accelerators, or on rotating devices with the subject positioned at some distance from the axis of rotation Under these conditions, there is a linear component of centrifugal acceleration which is proportional to the square of the angular velocity and the distance of the subject from the center of rotation Electrical stimulation of the vestibular system has been attempted but the inability to localize such stimulation has rendered this procedure of limited usefulness

Response. Indicators of the effects of vestibular stimulation which have been employed include a number of subjective effects such as the sensation of rotation, the appearance of rotary movements of the visual field, and the impression of tilting or change in orientation of the visual field Motor responses such as the nystagmus motions of the eyes in response to angular acceleration and counterrolling of the eyes in response to a change in orientation of the gravity vector have also proven useful. In addition, there may be gross motor reactions of the body, such as turning of the head or reorientation of the entire body or movements of the limbs in response to angular acceleration Responses within the nervous system have been studied by means of electrical recording from single units in various locations

The importance of a normal vestibular mechanism may be inferred by damaging the mechanism or destroying its function in experimental animals Labyrinthectomies may be produced surgically or the vestibular mechanism may be destroyed by drugs such as streptomycin with specific toxicity for this organ Human subjects with damage to the vestibular mechanism, either from pathology or from the use of toxic drugs have also been of interest in investigation of the system

Psychophysical Studies of Vestibular Function

THRESHOLD DETERMINATIONS

There is a minimum angular rate of acceleration which can be detected by a human subject Under optimum conditions, that rate is

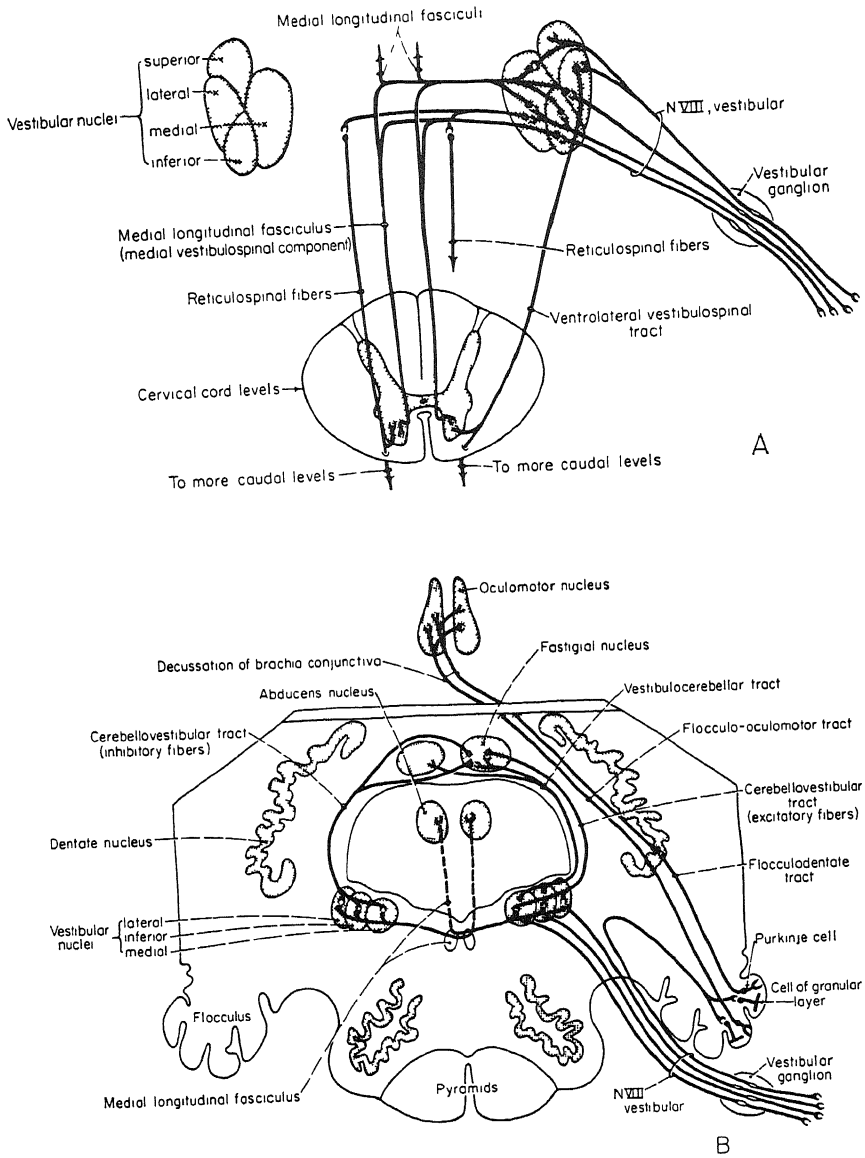


Fig 8 45 *A* Connections between the vestibular nuclei and the ventral horns of the spinal cord *B* Connections between the vestibular nuclei and the cerebellum (Crosby Humphrey and Lauer 1962)

probably of the order of 0 10°/sec ² (Clark, 1967, Clark and Steward, 1968) In order to achieve such low threshold values, it is necessary to employ a reference light point which rotates with the subject Apparently, vestibular imputs to the extraocular muscles in response to the low acceleration result in some eye movement and hence a relative movement of the image of the light spot on the retina which can be detected If subjects are rotated in complete darkness, thresholds of the order of 0 5°/sec ² or slightly lower are reported (De-Vries and Schierbeck, 1953, Meiry, 1966) It is not possible to find agreement on the question of whether certain pairs of canals may have lower thresholds than others (Clark and Steward, 1968). It is rather difficult to orient human observers for the selective stimulation of just one pair of canals

Lowenstein and Sand (1940) were unable to

find axes of rotation which did not produce any response from single fibers subserving the vertical canals in the ray One must be content with the selection of orientations of an axis of rotation with respect to the head which will maximize the stimulation of a given pair of canals Actually, for human subjects, there are no comparable data based even on such approximate maximization techniques for the threshold angular accelerations of selected pairs of canals

The best data available are probably those of Clark and Stewart (1968), which provide thresholds for 18 male subjects with their heads in each of three orthogonal positions with respect to an earth vertical axis of rotation Thresholds were determined with subjects in a seated position, upright, tilted 90° backwards, and on one side with respect to the surface of the rotating platform Results indicated mean thresholds of approximately 0 4°/sec ^{2} for the upright position and the backward tilted position The mean threshold for the side position was 0 67°/sec ^{2} The difference between this latter value and the others was not statistically significant Interaction between otolith excitation and semicircular canal excitation may well complicate the interpretation of results The thresholds reported are for the subjective detection of rotation in the absence of any visual cues whatsoever.

Threshold accelerations for the stimulation of nystagmus may be observed directly or recorded in terms of the electrical signal between two electrodes located on the temples. Such thresholds do not depend upon subjective report and this may be considered an advantage, however, the threshold accelerations for nystagmus are significantly higher than those for either the appearance of movement of a reference spot of light rotating with the observer or the subjective impression of rotation in the absence of any visual cues (Howard and Templeton, 1966)

The otolith organs, particularly the utricles upon which detection of head tilt is believed to depend, may be studied by tilting the body very slowly from an erect posture in a chair device to some position which deviates from the vertical If the rate of tilt is sufficiently slow to prevent stimulation of the semicircular canals, then the question may be asked how much change in orientation of the body with respect to the vertical can be detected by the subject A common procedure for studies of this sort is the presentation of an illuminated line of light, the orientation of which is controlled by the subject If the subject has been tilted to the right or left and is asked to position the illuminated line so that it is in a vertical orientation, then any deviation of the line from the true vertical provides a measure of the extent to which the observer is deceived with respect to the orientation of his own body Probably by reason of the orientation of the utricles within the head, the human subject is most sensitive to slight amounts of tilt away from the vertical As measured by the technique of orienting an illuminated line element, it has been demonstrated that for amounts of tilt of greater than 60° there is a tendency for subjects to underestimate the angle through which they have been tilted while for tilts of less than 60° there is an overestimation of the amount of tilt which has occurred (Miller and Graybiel, 1966) These effects were first reported many years ago (Muller, 1916)

Analogous with the nystagmus response to stimulation of the semicircular canals is the counterrolling of the eyes which occurs when the head is tilted away from the vertical The phenomenon does not lend itself readily to precise threshold determinations, however Counterrolling in the earth's gravitational field is approximately 10° for 60° of tilt away from the vertical No electrical index of counterrolling is readily recordable The dependence of counterrolling on otolith organs, which in turn are dependent upon a gravitational vector, has been demonstrated by studies of counterrolling in response to tilt in normal and labyrinthine defective subjects both on the earth's surface at $1 \times g$ and under zero gravity conditions during flight through a ballistic trajectory in an aircraft Labyrinthine defective subjects show little or no counterrolling under any circumstances Normal subjects show no counterrolling of the eyes with tilt of the body when they are in a zero gravity trajectory (Miller et al , 1966) At zero gravity the relatively high specific gravity of the otoconia is of no consequence for stimulation of the hair cells embedded in the macular substrate Change in the orientation of the body is not accompanied by an change in a linear acceleration vector

It is appropriate to assume that the

threshold angle of the otolith organs for tilt will vary with the magnitude of the prevailing gravitational acceleration force A functional relation between counterrolling in minutes of arc, and degrees of tilt at each of three levels of acceleration is illustrated in Figure 8 46 for both normal subjects and labyrinthine defective subjects (Miller and Graybiel, 1965) With increase in the gravitational force above the normal gravitational force at the earth's surface, there is an increase in counterrolling of the eyes directly proportional to the increased acceleration force up to 2 25 $\times$ g, the maximum studied, and probably beyond (Woellner and Graybiel, 1959) Graybiel and his colleagues, using a photographic superimposition technique, have been able to measure counterrolling to an accuracy of 5 min of arc

TEMPORAL DURATION OF EFFECTS OF VESTIBULAR STIMULATION

When a subject is exposed to constant angular acceleration about a given axis, a sensation of rotation is induced along with nystagmus movements of the eyes There appears to be an increase in the perceived angular velocity for approximately 30 sec followed by an impression of gradual decrease in angular velocity There is no decline in the nystagmus response to sustained angular acceleration which correlates with this subjective impression (Guedry, 1965) Thus, the attractive notion that the appearance of rotation of the visual world is somehow directly correlated with nystagmus movements of the eyes is probably in error Experiments have shown that the impression of rotation of the visual world may occur when there is no nystagmus detectable and that some nystagmus may be detectable when a subject is unaware of any subjective impression of rotation (Gray, 1960)

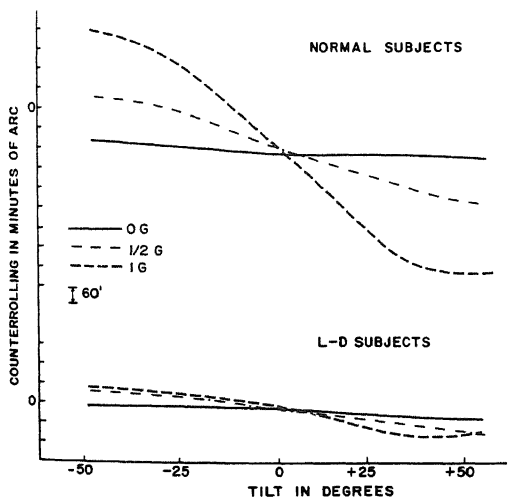


Fig 8 46 Mean counterrolling response of the eyes measured in minutes of arc as a result of tilting in angular degrees with respect to a resultant G-force Results are presented for both normal and labyrinthine defective (L-D) subjects (Miller and Graybiel 1965)

The effects of stimulation of the vestibular mechanism by angular acceleration may be measured in terms either of the impression of rotation, the so-called oculogyral illusion, or in terms of the measurable nystagmus movements of the eyes As suggested above, somewhat different results are obtained for these different indices Continuation of a nystagmus response with continuation of angular acceleration and reduction of the oculogyral illustion before termination of angular acceleration suggests that central factors may override the output of the peripheral sensing device as this influences conscious perception

For the usual exposures to angular acceleration for relatively short durations followed by the maintenance of a constant angular velocity for varying amounts of time, the perception of rotation and nystagmus both continue for 20 or 30 sec after angular acceleration has stopped in a normal subject This observation fits with the notion of the cupula as a damped torsion pendulum which gradually resumes its normal undeflected position when angular velocity is constant On the other hand, if rotation occurs about a horizontal axis, the perception of rotation continues for the entire duration of rotation even at constant angular velocity (Benson and Bodin, 1966) The explanation of this difference would seem to depend upon the fact that with rotation about a horizontal axis there is a continuously changing pattern of stimulation of the otolith organs This provides a continuing sensory input which corresponds to the continuing rotation even though the velocity is constant A curious aspect of this situation is that with rotation about a horizontal axis there is an immediate cessation of the impression of rotation when physical rotation is stopped In other words, although the semicircular canal stimulation resulting from deceleration should produce a continuing signal for some seconds after rotation is stopped, the perception of rotation stops immediately

Adaptation. As Adrian has pointed out (1943), the vestibular mechanism is not ideally suited to situations in which one is subjected to protracted angular motions Prior to the advent of high speed vehicles, particularly aircraft which can maneuver relatively freely in three-dimensional space, man was not ordinarily subjected to prolonged angular rotation The vestibular system may be ideally suited to coordination of the movements of the body and fixation of the eyes for short-term relatively low amplitude rotations, but this is clearly not the case for prolonged rotation After prolonged rotation at constant angular velocity, any deceleration results in sensory signals from the canals which are appropriate to rotations in the opposite direction and quite inappropriate to the actual motions involved This gives rise to difficulty of interpretation of the nature of the motion path of an aircraft in flight and has been blamed for innumerable aircraft accidents

The zero gravity environment of space is inconvenient from the standpoint of manipulating liquid foods and other aspects of daily living, and also presents certain physiological hazards with respect to venous return of blood to the heart and calcium metabolism For these reasons, it has been suggested that an artificial gravity might be useful for space stations and for space travel Such an artificial gravity could be achieved only by the maintenance of a constant angular velocity of the space station with occupants living at an appropriate radius so that the centrifugal force would provide an artificial gravity Problems arise by reason of the fact that the radius must necessarily be relatively short In moving about in such an environment, individuals will tilt their heads about various axes which may be orthogonal to the axis of rotation Such rotations give rise to Coriolis accelerations in which there may be a bilateral asymmetry of stimulation of the vestibular mechanisms The results of this effect can be most disturbing with respect to orientation and at worst may cause motion sickness

Investigations have been conducted in which volunteer subjects have lived in rotating environments on the earth's surface with the rotation about a vertical axis Free movement within such an environment is most disturbing and induces motion sickness in many subjects Adaptation to such an environment such that the effects of movement are no longer disturbing can occur if the rate of rotation is less than 6 or 7 r p m (Newsom et al , 1966) For rates of rotation above 10 r p m long-term adaptation is impossible (Graybiel, Clark, and Zarriello, 1960) Such a situation may actually be more disturbing than the situation which would exist in a rotating space station On the surface of the earth the major component of linear acceleration is the earth's gravitational component and a linear acceleration resulting from rotation is relatively small and constantly changing with movement of the individual along the radius of rotation

There appears to be little adaptation effect for the otolith organs when these are tilted to new positions in the earth's gravitational field Whereas the semicircular canals are stimulated only by angular acceleration and hence do not respond to constant angular velocity, the otolith organs are stimulated by the pattern of depression of the macular substrate by the otoconia Neural response, as described in greater detail below, continues relatively undiminished for each new position in which the head is placed If an observer is seated at a given radius on a centrifuge, rotating about a vertical axis, the otolith organs will respond to the resultant force of the centrifugal acceleration and the earth's gravitational field The result of this kind of stimulation is an impression of tilting of the visual world It is reoriented to conform to the direction of the resultant gravitational force as a new vertical reference This is known as the oculogravic illusion (Clark, 1967) On a centrifuge, the extent of tilt perceived by a subject is essentially constant for sustained acceleration at a constant angular velocity (Correra, Hixson, and Niven, 1968)

Habituation. The term, *habituation* is used as distinct from adaptation The former refers to effects of repeated exposure to angular rotation or accelerations Adaptation refers to effects which occur in a prolonged single exposure. It is well known that certain forms of habituation occur (Guedry, 1965) Tumblers and divers maintain a sense of their position in space while performing elaborate rotational movements which would be completely disorienting to anyone lacking in experience, if he could possibly perform the motions in the first place Figure skaters are capable of high speed rotation, as high as 280 r.p m for some sec-

onds, following which they can immediately go into another complicated skating routine or stand immobile They are able to suppress nystagmus motions of the eyes almost completely if they have an opportunity to look at their visual surroundings (Collins, 1966) Nystagmus movements have been detected by telemetry techniques in skaters, however, while they are in motion and after rotation if their eyes are closed It has also been demonstrated fairly conclusively that habituation is highly selective (Guedry, 1965) That is, a figure skater may be able to maintain complete control for rotation in a given direction with the head in an accustomed orientation On the other hand, if direction of rotation is reversed or if the head is tilted into an unusual posture, the result may be severe disorientation of the skater following the rotation

NYSTAGMUS

Nystagmus movements of the eyes usually associated with high angular acceleration consist of relatively slow components in the direction opposite to that of the acceleration and fast components in the direction of acceleration (Fig 8 47) Were it not extreme, the slow component would permit the maintenance of fixation on a point in the visual world in spite

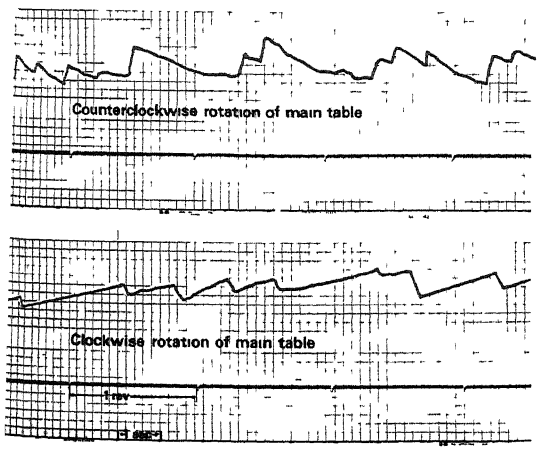


Fig 8 47 Records of horizontal nystagmus movements of the eyes of a cat to the right in the upper record and to the left in the bottom record These records appear typical of a nystagmus response to angular acceleration although they were obtained from an animal which was rotated on a small table located near the edge of a larger rotating table such that there was no resultant angular acceleration of the animal The experiment is described in the text (Money 1966)

of rotation Figure 8 48 illustrates the average slow phase nystagmus responses of 10 men subjected to an angular acceleration of 4°/sec ^{2} for a period of 45 sec (Crampton, 1965) It is evident that the direction of the slow phase is dependent upon the direction of rotation In addition, it is clear that the nystagmus effect continues for some 20 sec or more after the angular acceleration has stopped Cessation of the nystagmus in the initial direction is followed by a secondary nystagmus with the slow phase in the opposite direction The records in Figure 8 48 show some very slight evidence of *habituation* Average responses for the initial trials in each direction may be compared with average responses for the final trials in each direction The fact that the final trials show slightly lower maximum values is evidence of *habituation* This effect is even more marked with respect to the rate of reduction of nystagmus following termination of acceleration and the amplitude of the secondary nystagmus

There is also a suggestion of adaptation in the data presented in Figure 8 48 It appears that there may be a slight reduction in the nystagmus rate prior to the termination of acceleration The variability of the results is such that they cannot be interpreted as a clear-cut demonstration of adaptation, however As a matter of fact, nystagmus records, although they may show habituation, rarely show much if any adaptation Similar data for the same kind of exposure are presented in Figure 8 49 for the cat Here there is a much clearer demonstration of habituation, it is also somewhat clearer that there is no adaptation

For clinical tests using a rotating chair, a usual procedure is to rotate the subject at a rate of once every 2 sec for a total of 10 revolutions The chair is then stopped abruptly Nystagmus motions are observed which have been induced by the abrupt deceleration of the chair The slow phase is thus in the same direction as the original direction of chair rotation If, during rotation the head is nearly erect but tilted forward approximately 30°, then the nystagmus movement of the eyes will be from side to side or horizontal If the head is tilted down nearly 90° onto one shoulder with rotation about a vertical axis, then the direction of nystagmus will be up and down with respect to the head If the head is tilted back 60° or down 120° from the erect position,

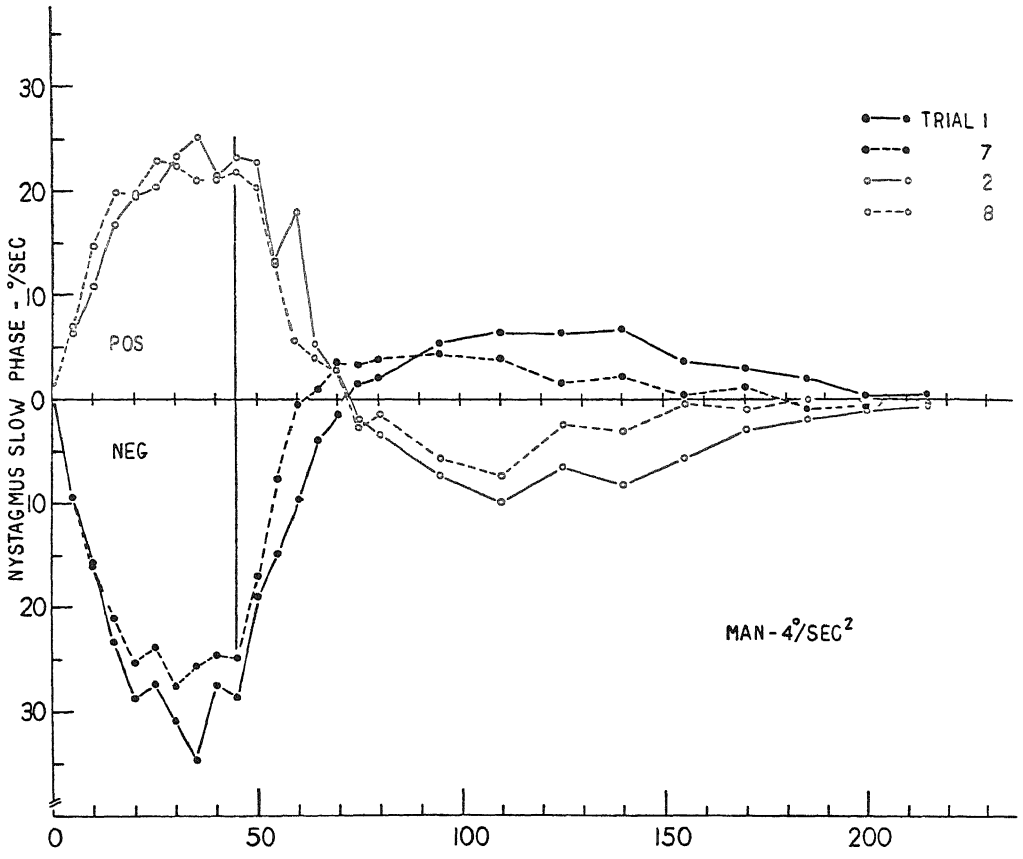


Fig 8 48 A record of the velocity of the slow-phase of horizontal nystagmus in man during and following exposure to angular acceleration at the rate of 4°/sec ² The termination of acceleration at approximately 45 sec is marked by a vertical line on the graph Results are the average responses for 10 men Results represented by solid circles are for counterclockwise or negative acceleration and results represented by open circles are for clockwise or positive accel eration The first trial for acceleration in a given direction is represented by a solid line the last trial by a dashed line The reversal of direction of nystagmus the so-called secondary nystagmus is shown in all of these records where they cross the zero point on the ordinate axis (Crampton 1965)

then the nystagmus will be rotary That is, rather than a linear motion there will be a torsional movement of the eyes This is slightly more difficult to observe but can readily be detected by referring to some unique characteristic of the iris or a vessel within the sclera A wide variety of nystagmus motions is possible dependent upon the position of the head with respect to the axis of rotation It is evident that the three pairs of semicircular canals work in concert to drive the appropriate extraocular muscles for compensation of motion of the head about any axis of rotation

Nystagmus may also be induced by the movement of objects in a given direction before the eyes. Thus, if the eyes are open during rotation, the *optokinetic* nystagmus will be in the same direction as the vestibular nystagmus and they will serve to enhance each other With the eyes open during deceleration, optokinetic nystagmus will tend to cancel out the vestibular nystagmus For this reason, in observing nystagmus effects with the rotating Barany chair, subjects are instructed to keep their eyes closed until the rotation of the chair has been stopped, at which point the eyes are opened for observation.

It is the slow phase of nystagmus which is initiated by the semicircular canals The quick component, which is merely a kind of "reset" motion, is attributed to unknown central pathways The characteristics of the slow phase provide us information as to vestibular response to stimulation

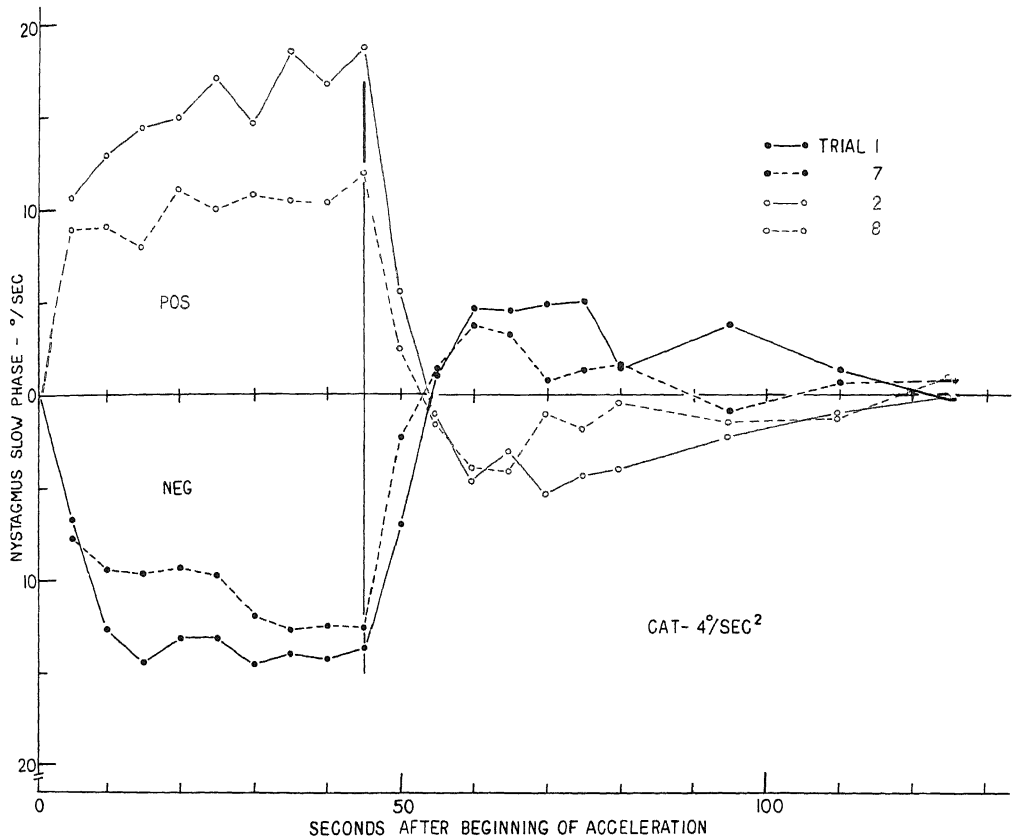


Fig 8 49 Results similar to those presented in Figure 8 48 representing average values for 10 cats the stimulus series was identical to that upon which the preceding figure was based (Crampton 1965)

Nystagmus has been observed in the absence of any angular acceleration of the semicircular canals If a small rotating table is located at the end of an arm which is rotating at the same angular rate but in the opposite direction, both about vertical axes, an animal on the table will maintain a constant geographic heading Under these circumstances nystagmus has been observed in man and in cats (Fig 8 47) The stimulus is a rotating linear acceleration vector which does not cause rotation of liquid in a closed vessel such as the semicircular canals Some action on the otolith organs might be inferred, but cats with defective otolith function also show the effect (Money, 1966) Nystagmus may result from effects of changing direction of linear acceleration on tissues of differing density within the ampullae of the canals

Orienting Responses. The vestibular stimulation which results in nystagmus movements of the eyes influences other musculature as well In a number of animals, nystagmus movements of the whole head may be observed Whole body orientation in the direction opposite to that of rotation may also occur There is a tendency for increased tone in the extensors on the side of the body in the direction of rotation and increased tone in the flexors on the side away from the direction of rotation The direction of these effects is, of course, reversed following deceleration Thus, if a subject is instructed to stand up and walk as soon as the rotation of the Bárány chair is stopped, he will tend to stagger or fall in the direction of the rotation

Electrophysiological Responses to Vestibular Stimulation

Electrical activity from single cells of the vestibular system has been a subject of study for a number of years (Ross, 1936, Lowenstein and Sand, 1936, Adrian, 1943)

SEMICIRCULAR CANALS

Gernandt (1949) reported on responses from three different classes of cells which he designated types I, II, and III The first of these responded by increased spike discharge with rotations in a direction which would cause bending of the cupula of the lateral canal in the direction of the utricle Rotation in the opposite direction resulted in an inhibition of the resting level of response of the unit The second type of cell responded with an increase in firing rate with acceleration in either direction The third type showed a decrease in firing rate with acceleration in either direction Gernandt had made these observations in the cat The first two types of cell were also found by Eckel (1954) and all three were reported by Duensing and Schaefer in 1958 The latter investigators also reported a new type of response, one in which the firing frequency of the unit was increased by rotation in a direction which would result in bending of the cupula of the horizontal canal away from the utricle and decreased for rotation in the opposite direction The response of this type was thus just the opposite of the first type of unit reported by Gernandt The first type may be thought of as conforming with predictions which one might make from the second law of Ewald concerning the appropriate direction of stimulation for the horizontal canals

There are a number of procedural factors which must be considered carefully in stimulating the vestibular system If the axis of rotation does not pass through the vestibular mechanism from which the recorded responses originate, then there will be a linear component of acceleration force imposed on the mechanism in addition to the angular acceleration This might result in stimulation of the otolith organs and the generation of responses which could be mistakenly attributed to semicircular canal function Another problem relates to the difficulty of so orienting the preparation that only one pair of canals is stimulated With the head erect, the lateral canals are stimulated most effectively, but there is some component of stimulation of the vertical canals as well This may be minimized or eliminated by inclination of the head forward by approximately 30°

Crampton (1965) has reported studies of single unit responses from the brain stem of the cat in which great care was taken to control the orientation of the head of the animal with respect to the axis of rotation In addition, precise control of the rate of angular acceleration was possible and the device permitted a smooth transition from rotation in one direction through 0 angular velocity to rotation in the other direction He found the type I cells of Gernandt which responded in accordance with Ewald's principle for lateral canal stimulation as well as the type of cell with an opposite response which had earlier seen reported by Duensing and Schaefer (1958) He did not find any cells which responded always with increase in firing rate or always with decrease in firing rate for either direction of angular acceleration He attributed the earlier reports of responses of this kind to possible slight displacement of the otolith organs from the axis of rotation The "non-Ewaldian" type of response was attributed by Crampton to a component of stimulation which influenced the vertical canals With the head erect, a direction of angular acceleration which would cause bending of the cupula in the horizontal canal in the direction of the utricle would also cause bending of the cupulae of the two vertical canals in the same direction However, as Ewald pointed out, the appropriate direction of the bending of the cupulae of the vertical canals for excitation is utriculofugal Therefore, these canals would provide an opposite effect as compared with the horizontal canals With the heads of his animals tilted downward 30° with respect to the stereotaxic horizontal plane, Crampton was unable to find any "non-Ewaldian" units With the head erect, they were found even when the contralateral branch of the eighth nerve had been sectioned, supporting his contention that they reflect ipsilateral effects in the vertical canals

Two different patterns of response were found in the cells studied by Crampton With positive acceleration there is a slight delay of the response which then builds up to a maximum value with approximately 45 sec of the beginning of angular acceleration With cessation of angular acceleration the response decays exponentially over a period of 50 sec or more Acceleration in the opposite direction results in suppression of response Approximately 20 sec after stimulation some activity reappears, but this then subsides below the normal resting level after about 100 sec and remains suppressed for over 5 min

Cells of the other sort respond with increase

in spike discharge frequency for angular acceleration in one direction and decrease in spike discharge frequency for angular acceleration in the other direction, but their level of activity does not return to a resting level with the cessation of angular acceleration The level of activity remains in the state induced by the angular acceleration for more than 5 min after acceleration exposure The level of activity may be "switched" to the lower or higher level, depending on its state, by accelerating the animal in the opposite direction These units thus appear to "remember" the direction of the last angular acceleration for a sustained interval after acceleration has stopped

OTOLITH ORGAN RESPONSE

Electrical activity of the otolith organs has been studied primarily in the fish by Lowenstein and his associates Recording directly from the otolith organs of the skate, Lowenstein and Roberts (1949) discovered several distinct types of receptor response patterns The stimulus consisted of tilting of the preparation in the earth's gravitational field In Figure 8 50, discharge frequencies are shown for a receptor in the utricle of the ray for each of a series of orientations of the animal through 360° of rotation The data were redetermined at each point with repositioning of the animal in the opposite direction Rate of change of position was not a matter of concern Recordings were made after the response had stabilized for a given position The unit which is represented in Figure 8 50 appears to be a position-sensing unit which provides a response frequency characteristic of a fixed orientation in space There is clearly some ambiguity in the response of such a unit, however Pairs of positions may be selected for which the impulse discharge frequency is the same, although the positions are quite different Other units have been found in the ray which respond maximally for a specific position such as upright, (Fig 8 51) and others have been found for which response frequency in a given position varies with the direction of rotation from which that position was achieved (Fig 8 52) Other receptors have been found which are obviously dependent upon rate of change of position for their response characteristics as well as position itself The combined function of all of these different receptor types would appear to provide a rich background of sensory input for the discrimination of the posture of the head in space as this changes in time

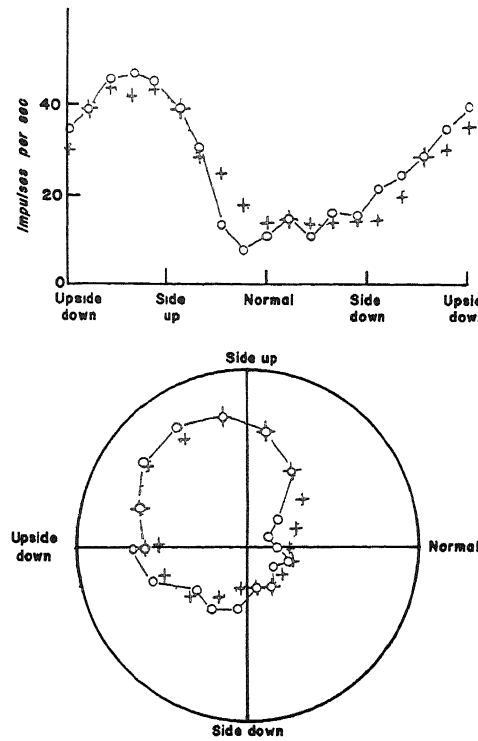


Fig 8 50 Rectilinear and polar plots of the firing frequency of a position receptor in the utricle of the ray as a function of skull position for full 360° rotation of the animal The continuous line represents clockwise rotation the dotted line represents counterclockwise rotation (Roberts 1967)

MOTION SICKNESS

The sensation of spinning of the visual world following rotation has been termed the oculogyral illusion This may not be an unpleasant experience and has been extremely useful as an index of the course of effects of vestibular stimulation The sensation of a whirling environment is also called vertigo It refers to the same phenomenon as the oculogyral illusion that is usually applied in extra experimental settings or under circumstances where the sensed rotations are disturbing The general feeling of *dizziness* may or may not include vertigo and is associated with unsteadiness and disequilibrium Vertigo may be a serious problem for the pilot of an aircraft if it results in confusion on his part with respect to his position or movement in space It is also serious as a concomitant of motion sickness

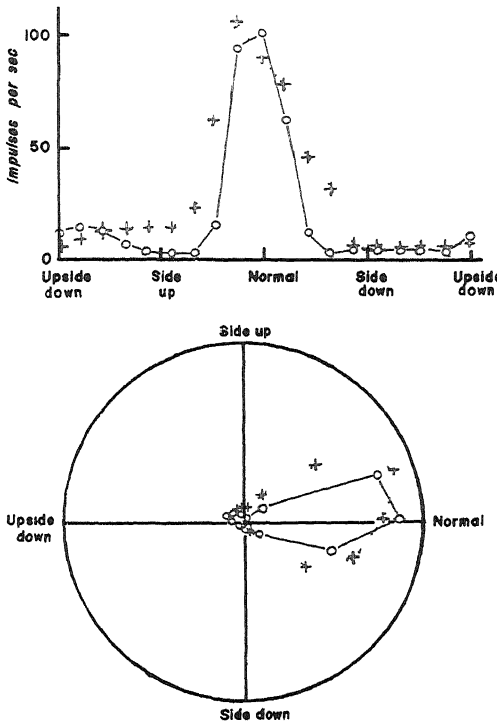


Fig 8 51 Rectilinear and polar plots of firing frequency as a function of skull position for a unit in the lagena of the ray which appears to respond only for positions in the vicinity of the normal upright position (Roberts 1967)

Unusual motion patterns in which the head is rotated about 2 axes simultaneously can produce motion sickness in anyone Some individuals are more resistant than others, but none is immune Motion sickness may also occur when visual and vestibular stimuli are not in accord Such a situation occurs for someone on board ship who is unable to see the stable horizon His visual environment moves with him and does not behave as would a stable visual world for the patterns of motion to which he is exposed by the pitching and rolling of the ship Motion sickness may also be induced when the body is perfectly stationary but the visual world, as seen in a motion picture film for example, changes as it might during airplane acrobatics or a roller coaster ride An important element in the induction of motion sickness is apparently lack of accord between visual and vestibular inputs The symptoms of motion sickness are similar to the effects of alcoholic intoxication An important component of each is the reaction of the autonomic nervous system This is immediately associated with the feeling of nausea, vomiting, and pallor Blood pressure may be lowered by as much as 10 mm of Hg and heart rate may be reduced by 8 or 10 beats/min

The nature of the motions encountered on shipboard are such that the cause of motion sickness has long been attributed primarily to otolith function (Sjoberg, 1968) The dominant role of the otoliths, however, has been seriously questioned (Graybiel and Miller, 1968) Since the discovery that the semicircular canals may respond to linear acceleration, particularly to change in the orientation of a linear acceleration vector (Miller and Graybiel, 1968), a more significant role has been attributed to the canals As mentioned above, motion sickness may be induced in the absence of any actual

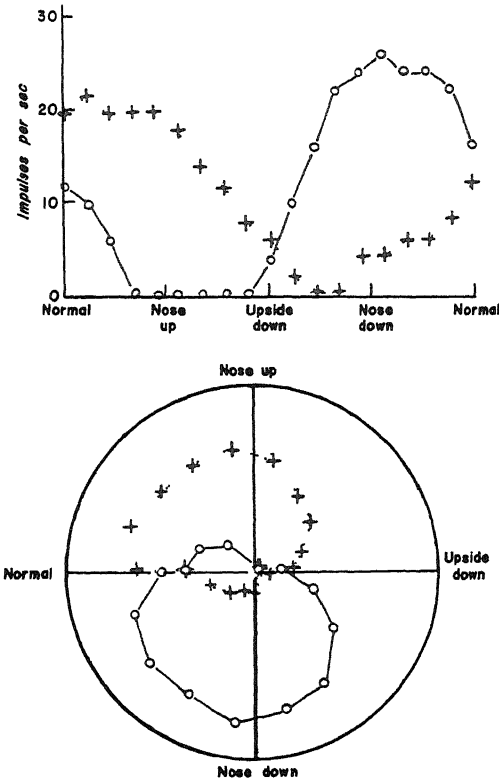


Fig 8 52 Rectilinear and polar plots of impulse frequency of a receptor in the utricle of the ray as a function of skull position during two 360° rotations in opposite directions It is evident that the response of the unit is influenced by direction of rotation Continuous curves represent clockwise rotation dotted curves represent counterclockwise rotation (Roberts 1967)

motion if visual stimulation depicting motion is presented (Miller and Goodson, 1961, Crampton and Young, 1953) Such visual motion stimulation is not effective in the induction of motion sickness in the absence of a functioning vestibular mechanism, however, (Crampton and Young, 1953) When visual inputs provide information that correlates with vestibular stimulation, vision may serve to reduce the incidence of motion sickness (Johnson and Taylor, 1961)

Inappropriate vestibular responses may occur following ingestion of alcohol, giving rise to the vertigo and dizziness experienced Alcohol does not increase nystagmus in the dark, however Increased nystagmus in the light following ingestion of alcohol is probably a result of the interference by alcohol with the normal visual suppression of nystagmus responses Nystagmus which occurs when the head is placed in certain positions after the ingestion of alcohol may be detectable for many hours after ingestion and after the level of alcohol in the blood has dropped to the normal range (Aschan, 1958)

An extensive survey of the effectiveness of a variety of drugs in combating motion sickness (Wood, 1968) strongly suggests that the effectiveness of such drugs is based on their action on the autonomic nervous system Both drugs which block parasympathetic activity and drugs which stimulate sympathetic activity are effective Scopolamine appears to be the most effective single drug The commonly used antihistamines are less effective and useful only in prevention of motion sickness under conditions of relatively "mild" stimulation

Labyrinthine Dysfunction

Any abrupt interference with the normal function of the labyrinth organs, particularly the vestibular mechanism, may be accompanied by the disturbing subjective symptoms of vertigo and dizziness along with ataxia and gross impairments of coordination A variety of conditions may be classified as labyrinthitis, accompanied by vertigo and the other symptoms referred to Inflammation of the vestibular branch of the eighth nerve, local infection within the vestibular mechanism and abnormal fluid balance within the system may be responsible for disturbances Treatment of such conditions is usually symptomatic Drugs which are effective in the prevention of motion sickness may be useful (see above)

Gross vestibular impairment accompanied by paroxysmal attacks of vertigo, subjective noise (tinnitus aurium) and a progressive impairment of hearing of the sensorineural type characterize a syndrome which is known as Meniere's disease Involvement of both vestibular function and the hearing mechanism suggests a localization of the disease process within the labyrinth It is generally attributed to disturbed fluid balance within the inner ear and distention of the cochlear duct has been observed upon postmortem examination of patients who had suffered from the disease Symptoms may be relieved by reducing the sodium intake Intractable cases have been treated by surgical destruction of the labyrinth, intracranial section of the vestibular division of the eighth nerve, or ultrasonic irradiation of the vestibular end organs for their destruction Labyrinthotomy performed under these circumstances has grossly different effects than those observed in animals The reason is probably that gross disturbance of the mechanism already exists before such intervention is carried out in man In a normal animal, destruction of the labyrinthine mechanism unilaterally has a result similar to excitation of the contralateral vestibular system Slow nystagmus movements are observed in the direction of the damaged side, there is a tendency for increased tonus of extensors on the undamaged side and of flexors on the damaged side

Vestibular function may be impaired by the administration of specific toxic substances such as streptomycin In doses of 2 g or more daily over a period of weeks, the occurrence of vestibular symptoms is fairly frequent Animal studies suggest a direct action on the peripheral sense receptors, although it has been suggested that in man the effect may be on the ventral cochlear nuclei The effects of streptomycin ingestion are usually bilateral If medication is stopped sufficiently early, the effects may be reversible Vestibular function is impaired before there is gross evidence of hearing loss Hearing impairments are greater for frequencies of 4000 Hz and above so that the deficit may not be noticed as these frequencies are of relatively little importance for the perception of speech

Diagnosis of vestibular impairment is accomplished by artificial stimulation of the vestibular mechanism. This may be accomplished by rotation in a chair or by caloric stimulation. Following rotation of the head about various axes, nystagmus movements of the eyes are timed and their duration is compared with normal standards. Impaired function is frequently associated with a greatly reduced duration of nystagmus. Similar observations are made following caloric stimulation. Caloric stimulation has the advantage of permitting selective unilateral stimulation. The selection of canals to be stimulated may be controlled to a limited extent by positioning of the head during stimulation. The effect of stimulation is believed to be based on convection currents so that the one canal which can be located in a horizontal plane will not receive stimulation.

LIGHT AND THE OPTICS OF THE EYE

Introduction · Light

Within the entire range of frequencies of electromagnetic radiation (70 octaves) only a narrow band of frequencies is effective in the stimulation of the eye It represents a range of only one octave, and in terms of wavelength extends from approximately 370 to 740 nm (10^{-9} meters) The distribution of frequencies and wavelengths of electromagnetic radiation is illustrated in Figure 8 53 Light is on the same continuum with γ rays, x-rays, and broadcast bands of radiation

The velocity of light in a vacuum is a universal constant It is the same for all frequencies When light enters any physical medium, however, its velocity is reduced The amount of reduction depends on the nature of the medium and the frequency of the light The ratio of the velocity of light in a vacuum to its velocity in a given medium defines the refractive index, n, of the medium The value of n for air is very nearly unity The higher the frequency, the larger is the value of n for a given medium The frequency of propagation remains constant independent of the medium through which light is transmitted and hence, as velocity is reduced, wavelength must also be shortened by a proportional amount

By application of the wave theory of transmission of electromagnetic energy, it can be demonstrated that light which passes from one medium into another will have its direction of propagation altered if it enters the new medium along a path which is other than perpendicular to the interface between the two media The ratio of the sine of the angle formed between the path of the light in the less dense medium and a perpendicular to the surface between the media to the sine of the angle formed by the path in the more dense medium with respect to the same perpendicular is equal to the index of refraction for the two media (Fig 8 54)

The principle of refraction is extremely important in a variety of ways If the surface between two media is not a plane but is curved, then rays of light which reach this surface at different points in a beam of large cross-section will be refracted by different amounts and will be propagated through the new medium in different directions It is on the basis of this principle that lenses may be formed from glass These can cause parallel beams of light to converge or diverge depending upon the design of the lens (Fig 8 55) Light of shorter wavelength is refracted a greater amount than light of longer wavelength Thus, a beam of white light which consists of a continuous distribution of energy for all visible frequencies from the longest to the shortest will be spread out if it enters glass at an angle other than perpendicular and the components of different frequencies will travel along different pathways This principle is turned to practical advantage by the use of the prism which permits the separation of the spectral components of light (Fig 8 56)

Although light is often specified in terms of its wavelength, it is more sensible to specify it in terms of frequency The reason for this is that the frequency of light is not altered when it passes from one medium into another of different refractive index Velocity does change and thus wavelength must also change a proportional amount For this reason, light is sometimes described in terms of its *wave number*, the reciprocal of wavelength in centimeters, which is directly proportional to frequency There are additional advantages in describing light in terms of its frequency or wave number associated with the characteristic form of the absorption spectra of certain pigments

QUANTUM THEORY OF LIGHT

Certain physical phenomena can only be understood in terms of the notion that light energy consists of a number of individual particles, each of which represents an indivisible,

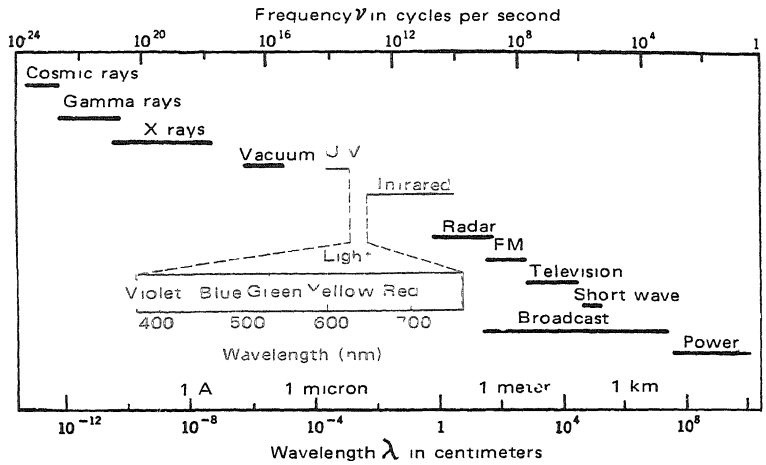


Fig 8 53 The electromagnetic radiation spectrum The region visible to the human eye is expanded (From McKinley 1947)

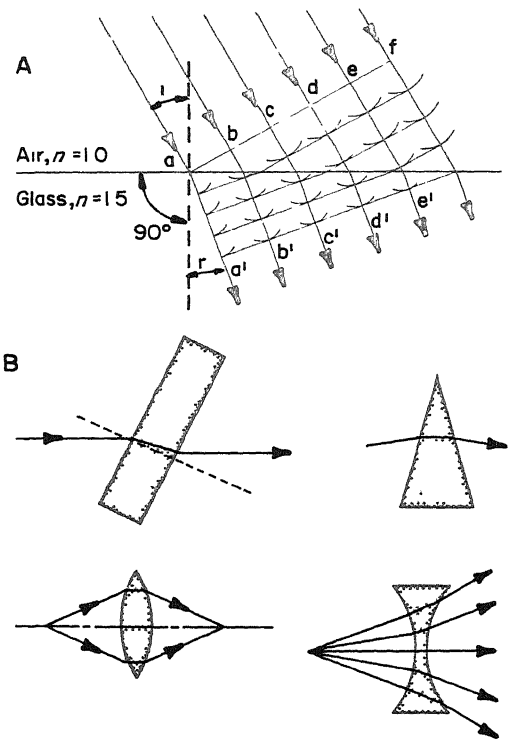


Fig 8 54 *A* refraction of light at a surface between glass and air Angle *i* is the angle of incidence and angle *r* is the angle of refraction Index of refraction *n* is given for each medium The rotation of wave front abcdef as it enters a medium of higher index of refraction at an angle other than normal to the interface is shown by wave front a'b'c'd'e'f' *B* deviations in the path of light as it is passed through glass of various geometric forms plain glass a prism a convex lens and a concave lens (Riggs 1965a)

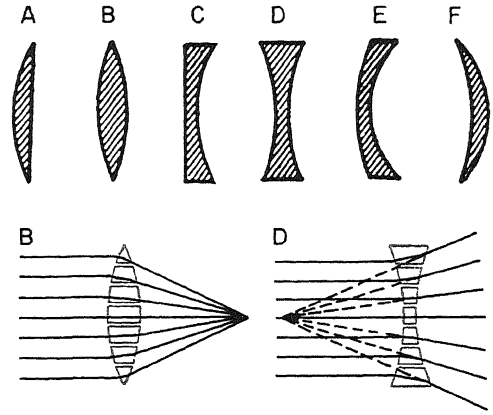


Fig 8 55 Cross sections of lenses *A* plano-convex *B* biconvex *C* plano-concave *D* biconcave *E* convexo-concave *F* concavo-convex The way in which refraction of a beam at curved surfaces may result in convergence or divergence is illustrated for *B* and *D*

minimum energy unit These are called *quanta* for the entire electromagnetic spectrum They are sometimes called *photons* for energy within the visible range The energy of a single quantum or photon of light is proportional to its frequency For a given frequency, the energy thus remains constant with changes in velocity as it passes through different media Energy may be computed from the following relation

$$E = h \times \nu$$

where E is the energy in ergs, h is Planck's

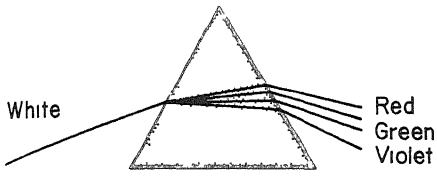


Fig 8 56 Dispersion of light by a prism

Constant, 6 6245 × 10^{-27} erg-sec and ν is frequency

THE MEASUREMENT OF LIGHT

Because only a narrow band of electromagnetic radiation is effective as a stimulus for vision, the measurement of light must take into account, not just the radiant emission of a light source independent of its wavelength, but also the effectiveness of the radiant energy as a stimulus to the eye Radiation of a continuously emitting source is measured in terms of energy per unit time For a point source, the measurement may be expressed in terms of the energy per unit time which irradiates a fixed area at a given distance from the source The fixed area at a given distance defines a solid angle and the radiation may thus be expressed in terms of ergs per second, or watts, per unit solid angle The rate of emission from the source is thus defined in terms which permit the computation of the amount of energy per unit time that will fall on any given surface at any given distance from the source The measure is one of the *intensity* of the source

If the source is extended rather than a point source, its emission must be expressed in terms of radiant energy per unit area Here, however, the orientation of the emitting surface with respect to the receiver or viewer becomes important This is taken into account by considering the area of the source projected on a plane normal to the axis along which observations are made The emission from an extended surface is termed its radiance and is expressed in terms of watts per unit solid angle per meter^2

Photometric Units. Even in the range of wavelengths to which the eye responds, the response is not uniform It depends upon the selective transmittance of the various optical elements of the eye itself and most importantly, on the absorption characteristics of the photopigments of the visual receptors Careful measurements have permitted the specification of a *spectral luminosity function* for human vision, V_λ (Fig 8 57) It is important to consider some of the limitations of this function In the first place, it represents the spectral response of the eye at "daylight" levels of illumination The spectral response of the eye changes at very low levels of illumination to the function represented by V'_λ These curves are average curves for a large number of observers and are based on the effectiveness of light as measured by a specific criterion The more effective a wavelength as a stimulus to the eye, the lower will be the radiant flux required for fusion of light flickering at a given frequency

The curves in Figure 8 57 are based on the reciprocals of radiant flux or energy requirements and have arbitrarily been set equal to unity at the wavelength of minimum energy requirement The low luminance V'_λ function actually represents a higher absolute level of sensitivity than the V_λ function

Unfortunately, the relation between the effectiveness of light energy based on criteria such as matching a standard or the apparent fusion of a flickering source is not perfectly correlated with the effectiveness of light energy for such tasks as reading and a variety of

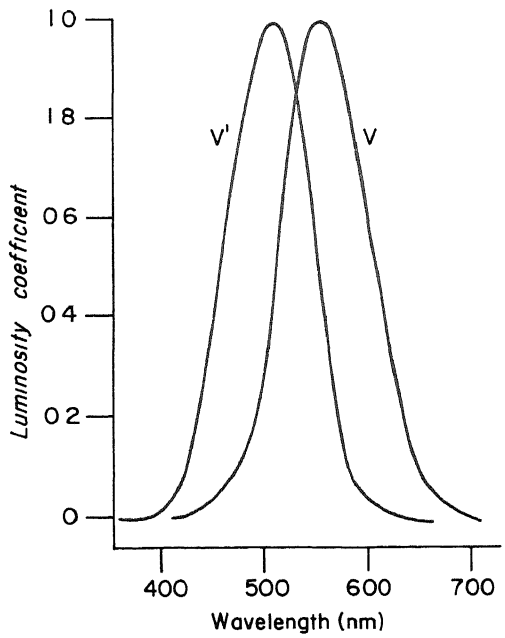


Fig 8 57 The scotopic (V') and photopic (V) curves of relative spectral luminosity as standardized by the Commission Internationale de L Eclairage (C I E) (From Wright 1958)

other visual functions of a practical sort Nonetheless, it does provide a reasonable approximation The units of light are based upon the weighting of a measurement of radiant power in appropriate units in accordance with the luminosity function presented in Figure 8 57 The photometric dimensional equivalent of the watt is the lumen At the wavelength where radiant energy is most effective in stimulating the eye, approximately 555 nm , one watt is equivalent to 680 lumens according to the accepted standard which provides a basis for photometric specification Radiant power of any other wavelength is less effective in accordance with the ordinate values of Figure 8 57 There is a variety of useful sources of detailed information on the specification of light for the interested student (Riggs, 1965a, LeGrand, 1968)

Optical Properties of the Eye

Prior to reaching the photosensitive layer of the retina, light which stimulates the eye must pass through the cornea, the anterior chamber, the vitreous humor, and a number of nonsensitive retinal layers (Fig 8,58) The distance from the corneal surface to the retina in the region of the foveal center of the eye is approximately 24 mm The optical elements of the eye must therefore provide a fairly high optical power in order to focus parallel rays which enter the eye at the retina The power of the eye is nearly 60 diopters Lens power is defined as the reciprocal of the focal length of the lens in meters The nodal point of the eye may be assumed to be approximately 17 mm in front of the retina and parallel rays may therefore be assumed to be brought to a point of focus in that distance The lens power of the eye is achieved primarily by the curvature of its corneal surface and secondarily by the power of the cystalline lens The refractive index of the cornea is 1 37, of the aqueous humor, 1 33, of the crystalline lens, 1 42, and of the vitreous body 1 33 Thus, maximum convergence of parallel rays occurs at the interface between the cornea with its refractive index of 1 37 and air with a refractive index 1 The radius of curvature of the cornea is nearly 8 mm at its anterior surface The refractive indices between the aqueous humor of the anterior chamber and the lens, and between the lens and the vitreous humor are sufficient to provide some additional positive power as a result of the convex shape of the lens The shape of the lens may be changed by the process of accommodation Contraction of circular fibers within the cilliary body relieves tension on the lens and its surface becomes more convex This permits an increase in the optical power of the eye by nearly 9 diopters Thus, accommodation is changed from a far point at infinity (for practical purposes 20 ft or more) to a near point of 10 or 11 cm

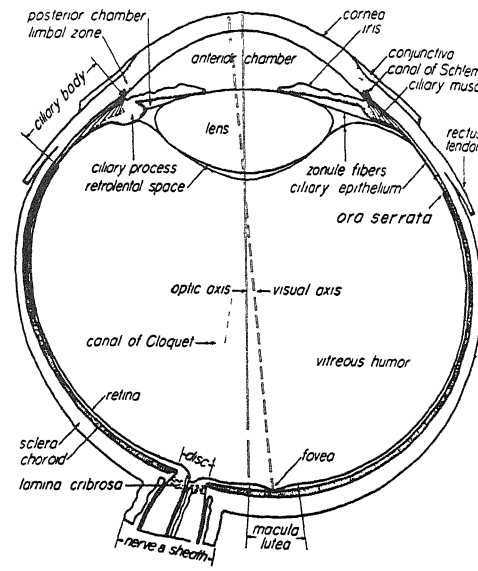


Fig 8 58 Horizontal section of the right human eye (From Walls 1942)

The intraocular media of the eye are not completely transparent The cornea begins to transmit light at wavelengths around 300 nm and its transmittance increases to a maximum in the region of approximately 1000 nm This is well beyond the practical limit for vision That limit is determined by the maximum wavelength which is absorbed by the photosensitive pigment of the retina Nonetheless, it is important to recognize that the cornea will transmit energy up to approximately 2000 nm , well into the infrared region The lens transmits energy over the same wavelength range but very little is transmitted at wavelengths longer than 1300 nm The aqueous humor absorbs virtually all energy at wavelengths longer than 2700 nm and the vitreous humor absorbs virtually all energy at wavelengths longer than 1600 nm The minimum wavelength effective in stimulating vision is of the order of 370 nm in the average adult

Shorter wavelengths may stimulate vision directly in the very young but there is a gradual yellowing of the ocular media with increasing age and accompanying attenuation of near ultraviolet energy down to 300 nm Between 350 and 400 nm , radiant energy may cause fluorescence of the lens As indicated, the fluorescent energy will be lower than the energy of the radiation which causes it and hence will be at longer wavelengths The appearance of light when the eye is stimulated by radiant energy at 350 nm may therefore result from longer wavelength fluorescence and cannot be interpreted to prove that the eye is capable of detecting energy at a wavelength of 350 nm

THE PUPIL

The total amount of light which may enter the eye is limited by the iris diaphragm or pupil This may vary in diameter from approximately 2 to 8 mm This four to one change in diameter corresponds to a 16 to 1 change in area and a commensurate change in the level of retinal illumination for a given level of illumination at the cornea A sphincter muscle innervated by the parasympathetic nervous system causes contraction of the pupil, a dilator muscle, innervated by the sympathetic nervous system causes dilation of the pupil Any abrupt increase in the level of illumination causes a pupillary reflex, observable in both eyes even though only one eye may be illuminated The diameter of the pupil is stopped down to a smaller size and then, if the illumination remains the same level, the pupil will open slightly to a steady state condition a little larger than the maximum initial constricted size Reduction in pupil size affords some protection for the eye at high levels of illumination, but this may not be its most important function

Almost any optical system suffers from certain imperfections Variation in the refractive index with the wavelength of light results in a different lens power for different wavelengths and therefore some distortion at the image for any illumination which represents a broad band of wavelengths In addition, the eye suffers from spherical aberration That is, rays of light which enter the lens near its periphery may be brought to a focus at a slightly different depth than other rays entering near the center of the lens These two types of aberration are illustrated in Figure 8 59 The smaller

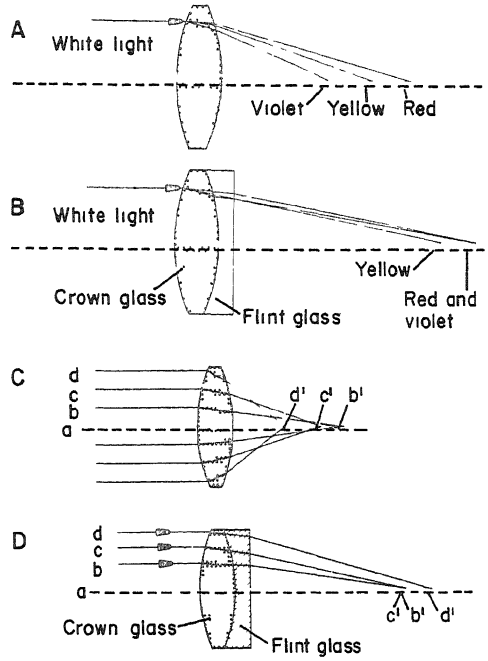


Fig 8 59 Chromatic and spherical aberration *a* chromatic aberration in a simple lens *b* achromatic lens partially correcting chromatic aberration *c* spherical aberration in a simple lens *d* achromatic lens partially correcting spherical aberration (Riggs 1965a)

the pupil of the eye, the less will be the degradation of the retinal image by reason of spherical aberration In addition, the greater will be the depth of focus of the eye as an optical system and hence the less will be the distortion due to chromatic aberration Therefore, the higher the level of illumination and, consequently, the smaller the pupil, the sharper will be the retinal image

The accommodative reflex, adjustment of lens size to decrease the focal length of the eye, is accompanied by pupillary constriction independent of any change in level of illumination The reduced pupil size increases depth of focus and thus provides a sharper retinal image when objects are viewed close to the eyes Unfortunately, there is a limit to the advantage which may be gained by pupillary constriction When light passes through a very small aperture, diffraction effects occur These may be viewed as interference and reinforcement effects such that the image of a point of light will consist of a central bright spot surrounded by a series of dark and light rings Thus, if the pupil is too small, diffraction effects will degrade the image An optimum

pupil diameter is probably between 2 and 3 mm For larger pupil diameters, spherical and chromatic aberration effects degrade the image, for smaller pupil diameters diffraction effects degrade the image

THE STILES-CRAWFORD EFFECT

Under normal conditions of daylight viewing, a specialized set of receptors, the cones, provide a basis for visual sensitivity Apparently, the geometry of these receptors and their orientation within the retina is such that they absorb light more efficiently the more nearly its point of entry into the eye is on the central visual axis of the eye Deviation of the point of entry away from the center by as much as 4 mm with the pupil at maximum diameter may reduce the effectiveness of light as a stimulus for the cones by a factor of 10 The relation of the luminous efficiency to eccentricity of the point of entry is illustrated in Figure 8 60 This indicates that the gain in retinal illumination in terms of actual effectiveness of stimulation of cones is not directly proportional to the increase in pupil area with dilation of the pupil No effect of this sort is found at low levels of illumination when the scotopic receptor system, dependent upon the rods, provides a basis for visual sensitivity Apparently the geometry and orientation of the rods are not as critical in determining their ability to absorb incident light

CHANGES IN THE LENS WITH AGING

It has been indicated above that the lens is capable of effecting a change in optical power of the eye of as much as 9 or 10 diopters Unfortunately, this does not remain the case indefinitely Almost from birth there is a gradual stiffening of the lens such that the range of change in curvature is gradually reduced with age Beginning at an age somewhere in the vicinity of 45 years, most people must begin to wear glasses for reading Their eyes are no longer capable of providing good accommodation for printed material at a comfortable reading distance If the material is moved sufficiently far away to provide for a sharp image on the retina, the size of the retinal image is too small for resolution of the letters The condition is easily rectified by the use of a positive lens in front of the eye

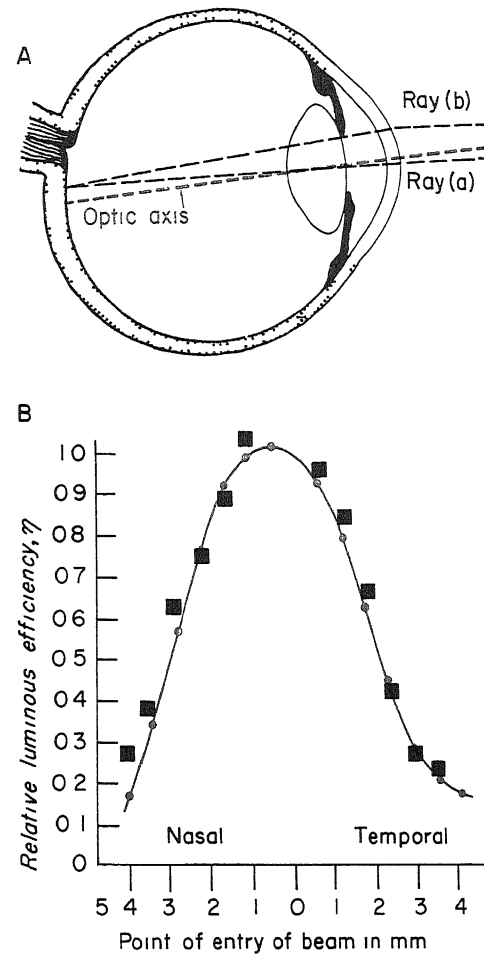


Fig 8 60 *A* the eye in horizontal section to show the Stiles-Crawford effect *B* the relative luminous efficiency of light entering the pupil at various points in a horizontal plane relative to the center of the eye Results for the left eye are indicated by squares for the right eye by dots (From Stiles and Crawford 1933)

OPTICAL DEFECTS

In a normal eye with relaxed accommodation, objects at an optical distance of infinity should be brought to focus on the retina Such an eye is said to be emmetropic If the point of focus is located out in front of the retina it is necessary to move the object toward the eye, displacing the image further to the rear of the optical elements of the eye, until it falls on the retina, in order to see it sharply Such an eye is said to be myopic If the image of an object at optical infinity falls behind the retina, then increased accommodation of the lens is neces-

sary in order to bring it forward onto the retinal surface Such an eye is said to by hypermetropic

Myopia is usually attributed to an elongation of the eyeball, although it may well involve abnormalities of curvature of the optical elements Conversely, hypermetropia is attributed to a foreshortening of the axial length of the eye ball It, too, may involve deviations from normal of the optical elements, however Myopia is correctible by the use of a negative or diverging lens in front of the eye in order to permit distance vision Frequently a myope without too severe a deficit may be able to read perfectly well without correction The hypermetrope will require a positive lens for reading, or, at least, will have to start wearing glasses at an earlier than normal age as his accommodative reserve is lost (Fig 8 61)

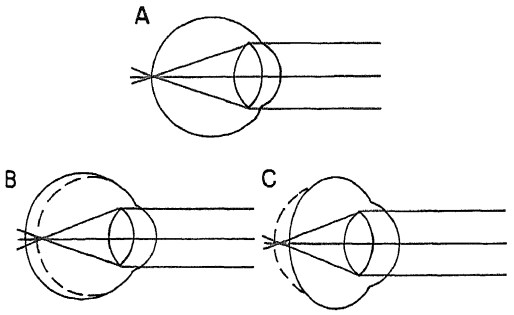


Fig 8 61 Comparison of shapes of the eyeball which may be associated with *A* emmetropia *B* myopia *C* hypermetropia

VARIATION IN THE SPATIAL RESOLUTION CAPABILITY OF THE EYE AS A FUNCTION OF THE LOCATION OF THE STIMULUS

Unlike a photographic film, the resolving power of the photosensitive layer of the eye, the retina, is not uniform over its entire surface A condition of optimum clarity is achieved for images which fall on or in the neighborhood of the visual center of the eye, the fovea (see Fig 8 58) The distribution of receptors over the retina is nonuniform with the greatest concentration of cones in the foveal center The concentration of rods increases from a point just outside the rod-free fovea to a maximum at a distance of some 20° from the foveal center (Fig 8 62) The distribution of receptor

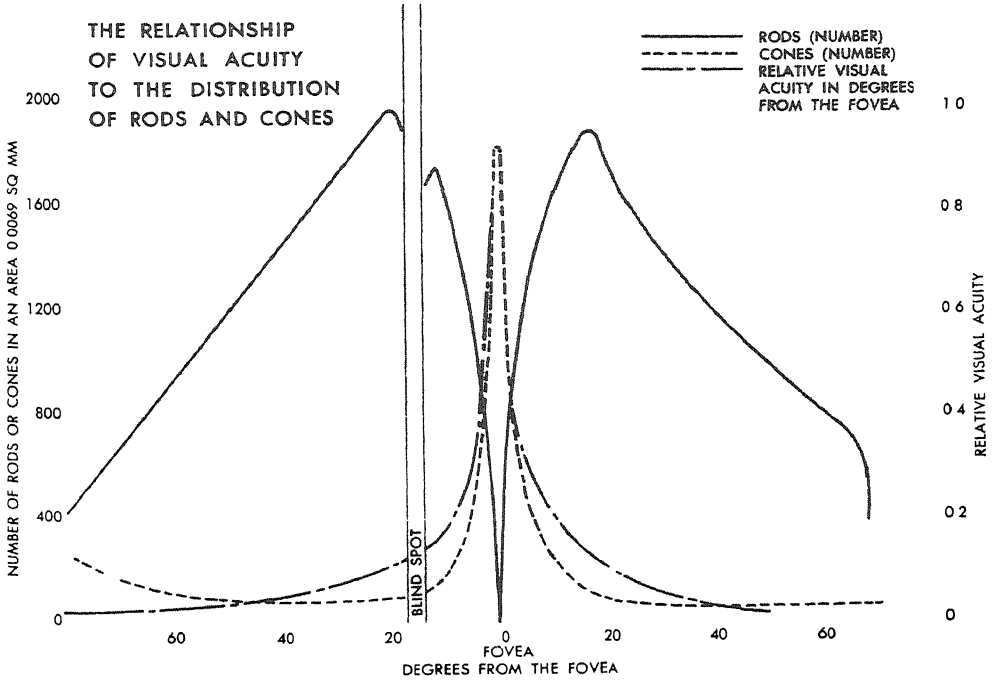


Fig 8 62 Distribution of rods and cones along a horizontal meridian Parallel vertical lines show the limits of the blind spot Visual acuity for a high luminance as a function of retinal location is included for comparison (From Woodson 1954 data from Osterberg 1935 and Wertheim, 1894)

density is discussed in more detail in the following section In addition to the nonuniform distribution of receptors, there is a region in the normal eye which is devoid of any receptors That is the so-called optic disk (Fig 8 58), where nerve fibers leave the eye to enter the optic nerve and where the retinal circulation enters and leaves the eye This is a region of 3° or more in diameter centered at a point 15° to 17° from the foveal center and with its center just slightly below the nasal horizontal meridian The presence of the blind spot is usually not noticed, even for monocular viewing The reason for this is the fact that it is not associated with a dark region in the visual field but is seen as identical with the surrounding region It is as if the surrounding region "fills in" the blind area in conscious perception The blind spot may readily be demonstrated as shown in Figure 8 63 If the left eye is covered and the small cross is fixated, the black dot will disappear as the page is brought to within approximately 14 inches from the eye It will remain invisible until the cross is to within approximately 10 inches of the eye The vertical extent of the blind spot may be explored by scanning the page vertically along an imaginary line through the fixation cross with the page held at a distance of approximately 12 inches

Information on the optical properties of the eye, at a variety of levels of detail, may be found in a number of recent texts (Brown, 1965a, Le Grand, 1968, Davson, 1962)

THE ANATOMY OF THE VISUAL SYSTEM

The Eye

The eye is enclosed by three membranes, the outer of which consists of the opaque sclera over most of the globe which is continuous with the transparent cornea of the anterior surface of the eye The major blood supply of the eye lies in the heavily pigmented choroid coat directly below the sclera The pigmentation of this layer serves to absorb light within the eye, preventing back scatter, and to reduce entry into the eye except through the cornea and lens At the anterior pole of the eye the choroid coat is continuous with the iris diaphragm, or pupil, and the ciliary body The third, innermost layer of the eye is the retina itself Visual receptors, neural connections of these receptors and additional vascular tissue lie in the retina

THE CORNEA

The cornea is a highly transparent tissue, devoid of blood vessels but well supplied with undifferentiated nerve endings (Fig 8 64) It is nearly 12 mm in diameter, less than a millimeter thick at the center and approximately 1 mm thick at its edges The radius of curvature of its anterior surface is about 7 8 mm , that of its inner surface is approximately 6 8 mm Detailed examination of the structure of the cornea reveals that it is made up of five distinct layers The outermost is continuous with the *conjunctiva*, a membranous covering of the anterior surface of the eye which is folded back on the inside of the eyelids Below the epithelial conjunctive is an elastic layer, *Bowman's membrane* Most of the thickness of the cornea is made up of the *substantia propria* which constitutes nearly 90% of the 1 mm of thickness of this tissue The *substantia propria* itself consists of approximately 60 collageneous layers made up of parallel rows of fibers separated by cell spaces Posterior to the *sub-*

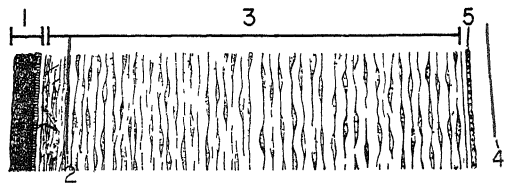


Fig 8 64 Vertical section of human cornea (lying horizontally) *1* corneal epithelium *2* anterior elastic lamina *3* substantia propria *4* posterior elastic lamina *5* endothelium of the anterior chamber



Fig 8 63 Demonstration of the blind spot With the left eye closed fixate the small cross Vary the distance of the page until the black spot disappears The horizontal extent of the blind spot is indicated indirectly by the range of depth in which the black spot is invisible Some indication of the vertical extent of the blind spot may be obtained by scanning the page along an imaginary vertical line through the fixation cross

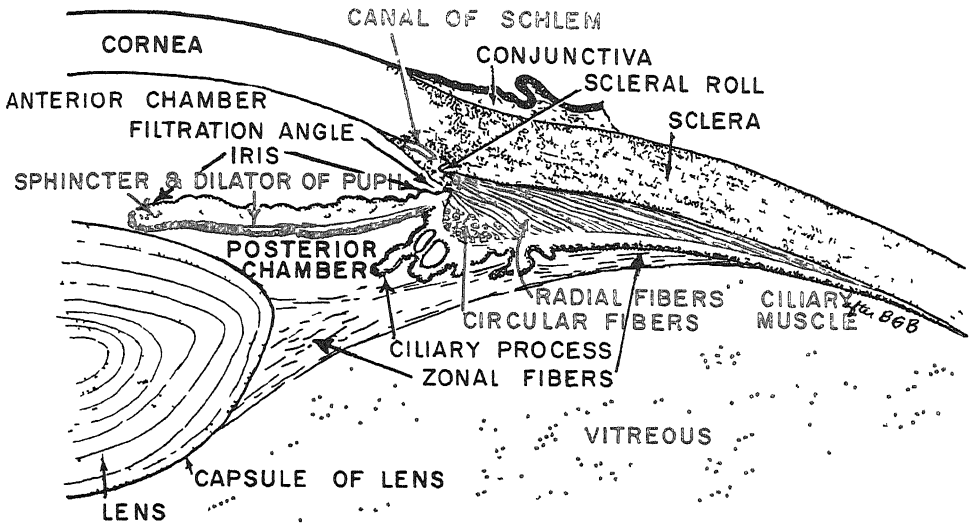


Fig 8 65 Detail of the anterior segment of the human eye (Redrawn from Weymouth 1955)

stantia propria is another elastic layer, *Descemet's membrane*, 5 to 7 μ in thickness The fifth and inner layer of the cornea is the endothelium comprised of flat cells which are in contact with the aqueous humor Although it is without a blood supply, the cornea consumes oxygen at a relatively high rate as a result primarily of epithelial activity (Cogan and Kinsey, 1942, Herrmann and Hickman, 1948)

Although only free nerve endings have been found in the cornea, there is evidence that it may be sensitive to pressure (Rowbotham, 1939), touch, warmth and cold (Lele and Weddell, 1956) The ability of free nerve endings to serve a variety of sensations in other parts of the body is well established (chap 7)

AQUEOUS HUMOR

The transparent fluid which fills the anterior and posterior chambers on either side of the iris diaphragm is formed from blood plasma The blood supply is that to the ciliary body (Davson, 1950) Production of aqueous humor occurs continuously and there must be a route for its drainage from the eye That route is the canal of Schlemm (see Fig 8 65) A complete turnover of aqueous humor occurs approximately once every hour If for any reason the outflow is blocked, continuing production of fluid may cause an increased intraocular pressure Normally, the intraocular pressure is approximately 15 to 18 mm of Hg higher than the intracranial pressure (Adler, 1959) This pressure differential helps to maintain the shape of the eye and, consequently, the appropriate spacing of its optical components

Aqueous humor is similar to blood plasma but contains a lower concentration of protein Metabolic requirements of the crystalline lens and of the cornea are supplied in part by the aqueous humor

THE PUPIL

The relation of the pupil or iris diaphragm to the lens of the eye is portrayed in Figure 8 58 The iris diaphragm is a highly pigmented tissue which arises from the anterior surface of the ciliary body (Fig 8 65) Coloration of the iris results from the nature of the pigment in the front surface of the iris Pigment at the posterior surface is black and helps to reduce light scattering as does the choroid coat.

The sphincter muscle which permits contraction of the pupil forms a ring approximately 0 8 mm in thickness around the pupil It can shorten by nearly 80% of its maximum length in constricting the pupil As mentioned above, its innervation comes from the parasympathetic nervous system The dilator muscles are radial fibers of myoepithelium within the iris These are innervated by the sympathetic nervous system (Adler, 1959) When a bright light is shined in the eye, there is a latency which may be as long as half a second prior to the reflex pupillary contraction The neural pathways which control this reflex must, therefore, involve several neurons Fibers which mediate the response leave the optic tract,

cross the medial border of the lateral geniculate body, and project toward but do not enter the superior colliculus They then travel medially and rostrally to the pretectal region From here, impulses are relayed to the oculomotor nuclei bilaterally and then to the ciliary ganglion (Duke-Elder, 1961) The exact pathway for reflex dilation of the pupil in darkness is not known It does not include the pretectal nucleus The ciliospinal center in the upper thoracic level of the cord is involved and fibers from there pass to the superior cervical ganglion of the sympathetic system Postganglionic fibers innervate the dilator muscles

THE LENS

The lens is a biconvex body which contains neither vessels nor nerves (Figs 8 58 and 8 65) It is made up of many concentric layers of fibrous cells, is approximately 65% water and 6% fat, and contains more protein than any other tissue The lower refractive power of the lens cortex as compared with that of the cornea provides some correction for both chromatic and spherical aberration of the eye The yellowing of the dioptric elements of the eye with age is due primarily to the lens itself (McEwen, 1959) It absorbs nearly 10% of the visible light which enters the eye with the level of absorption being highest at the shorter wavelengths (Weale, 1954) Infrared and ultraviolet light absorbed by lens protein may cause damage Energy for lens metabolism is derived from glucose oxidation Oxygen, ascorbic acid and glutathione are derived from the aqueous humor (Davson, 1949)

The lens is held in place by an elastic capsule, the ciliary zonule (Adler, 1959) (Fig 8 65) With the muscles of accommodation at rest, there is sufficient tension on the capsule to maintain the lens somewhat flattened Contraction of the ciliary fibers through the oculomotor and short ciliary nerves under the control of the parasympathetic system reduces the tension on the lens with a resulting increase in its convexity This permits accommodation for near objects Stimulation of radial fibers via the long ciliary nerves mediated by the cervical sympathetic system can result in accommodation change of as much as 1 5 diopters (Olmstead, 1944) Accommodation to near objects is more rapid than accommodation to distant objects

VITREOUS HUMOR

The vitreous humor is a gel-like substance containing a network of thin fibers of a highly hygroscopic vitrein protein similar to gelatin (Duke-Elder, 1937) With the aid of an electron microscope, it is possible to distinguish the fibrils and interfibrillar substance of the vitreous body of the human eye Fibrils are approximately 67 Å (angstroms) and deviate remarkably little from this size Spherical particles of 100 Å in diameter have also been found (Schwarz, 1961)

THE RETINA

During ontogenetic development, the eye is formed from optic vesicles that arise within the neural groove The optic stalks arise out of the neural tube between the diencephalon and the telencephalon Eyes are formed when the ends of the stalks contact the surface ectoderm The retina is derived out of ectoderm of the neural tube itself It thus has its origin directly out of the central nervous system and its cells may be considered a peripheral extension of the central nervous system Cell development of the retina occurs in stages during the first of which cells arise which migrate outward, eventually to become ganglion cells "Outward" during this development stage is actually "inward" with respect to the eyeball itself as it is finally formed In the next stage of development, the bipolar cells are differentiated, and finally, in the innermost portion of the developing tissue, one of the outer layers of the developed retina, the receptor cells themselves develop

The most prominent aspect of the appearance of the retina is its vascular tissue No vasculature intrudes into the central fovea, that region of the eye which is optimally organized for detailed vision The various retinal layers are thinner in this region (Fig 8 66), ganglion cell bodies are displaced laterally away from the foveal depression and the elongated cones of the central fovea occupy a substantial portion of the retina's thickness In their course toward the optic disk, fibers are deviated around the fovea There is thus a minimum of tissue between the receptor cells and the vitreous humor in the foveal center of the retina (Fig 8 67)

The ten layers of the retina which are frequently discriminated are illustrated schemati-

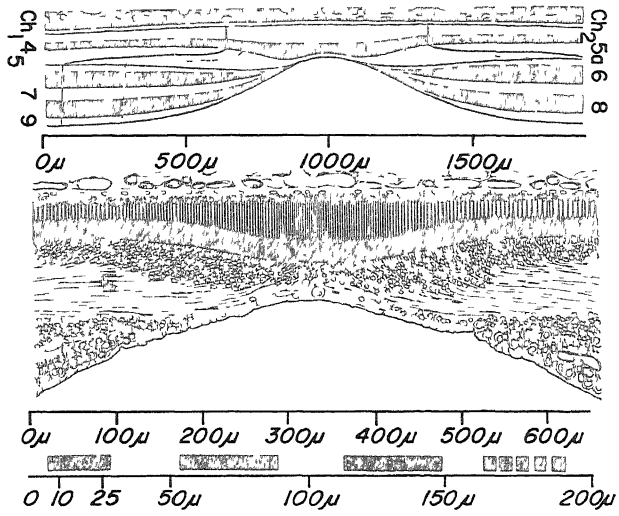


Fig 8 66 Central fovea of the human retina The upper part of the figure is a diagramatic illustration of the arrangement of the layers in the central region Ch choroid coat Numbered layers correspond with the retinal layers as described in the text (From Polyak 1957)

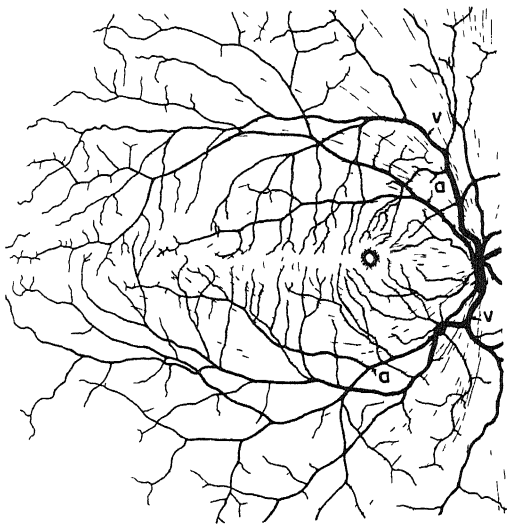


Fig 8 67 Distribution of optic nerve fibers (thin lines) and blood vessels (arteries a veins v) in the retina of an adult *Rhesus macaque* The disc of the optic nerve is at the right margin of the figure The central fovea free of vessels appears to the left of the disc (From Polyak 1941)

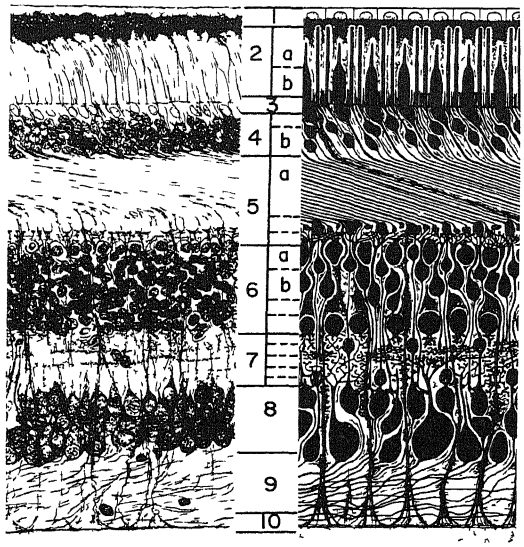


Fig 8 68 Vertical sections of the human retina from the central area midway between the areal periphery and the central fovea Right and left halves of the figure represent different methods of staining The various layers are described in the text (From Polyak 1957)

cally in Figure 8 68 on the basis of microscopic studies with different types of staining Light first enters the retina at layer 10, the inner limiting membrane It then passes successively through the layer of optic nerve fibers, the ganglion cell layer, the inner plexiform layer and the inner nuclear layer The latter contains the cell bodies of bipolar cells and horizontal cells The vessels which supply blood to the tissue of these last three layers are contained in the layers themselves These are the vessels visible in Figure 8 67 The outer plexiform layer contains the fibers from the primary receptors and their synapses with bipolar

and horizontal cells The outer nuclear layer contains the cell bodies of the receptors, the outer limiting membrane separates the region containing the cell bodies of the receptors from their outer segments The second layer is comprised of the outer segments themselves of the rods and the cones The final layer is the heavily pigmented epithelial layer Because light must pass through most of the other layers of the retina before it reaches the primary receptors, the eye of vertebrates is referred to as an inverted eye

It is not possible to say precisely how many receptors are present in the retina of the human eye but frequently quoted figures are 120,000,000 rods and 5,000,000 to 7,000,000 cones (Davson, 1949) Relative distributions of rods and cones are illustrated in Figure 8 62 along with a function which illustrates the visual acuity or spatial resolution capability of the eye, the latter is closely related to the distribution of cones There are no rods in the central foveal region of the retina This region is occupied exclusively by cones, and their number in this rod-free region is somewhere in the neighborhood of 30,000 to 35,000 The angle subtended by the rod-free region at the nodal point 17 mm in front of it is approximately 1° It is convenient to refer to locations on the retina in terms of visual angle rather than actual distance Such a scheme makes it easier to relate the position of an image on the retina to the position of objects in the visual world relative to the point of fixation of vision in the visual world The geometry of the eye is such that a visual angle of 12 sec of arc subtended at the nodal point represents a distance measured along the retina of approximately 1 μ

There are approximately 1,000,000 ganglion cells in the retina, the axons of which comprise the 1,000,000 fibers of the optic nerve It is therefore evident that many receptors must converge on single fibers in transmitting their activity centrally in the visual system There may well be individual ganglion cell pathways available to individual cone receptors of the central fovea, but it is well established that there are many cross connections among receptors within the retina, even in the central fovea

THE RETINAL RECEPTORS

The appearance of rods and two types of human cones based on light microscopy is illustrated in Figure 8 69 The foveal cones are perhaps even more rod-like in appearance than the rods themselves The development of electron microscopy has disclosed many anatomical details of these receptors (Fig 8 70) The receptors in this figure are those of *Necturus*, the mud puppy It is clear that the disk-like structures in the outer segments of the rods differ from those in cones The discs in the rod outer segments appear to be completely separated while those in cones actually represent repeated infoldings of the plasma membrane on the side of the outer segment opposite the ciliary fibrils Primate cones, however, appear to be formed differently (Cohen, 1961, Dowling, 1965) The layers of the outer segments of cones are individual flattened discs just as in rods except that they lack a characteristic rim structure seen in the rods (Fig 8 71) There are no continuities with the plasma membrane in

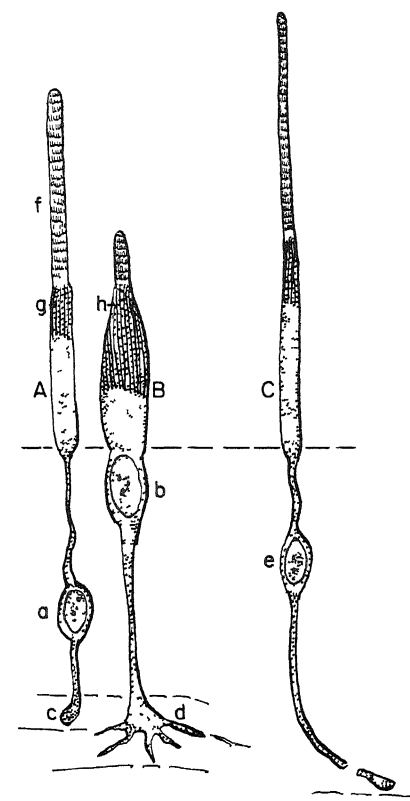


Fig 8 69 Drawings of human rods and cones as seen under light microscopy *A* rod *a* nucleus *c* central connections *g* ellipsoid *f* outer segment *B* cone *b* nucleus *d*, central pedicle *h* ellipsoid *C* cones in the fovea centralis *e* nucleus (Cajal 1904)

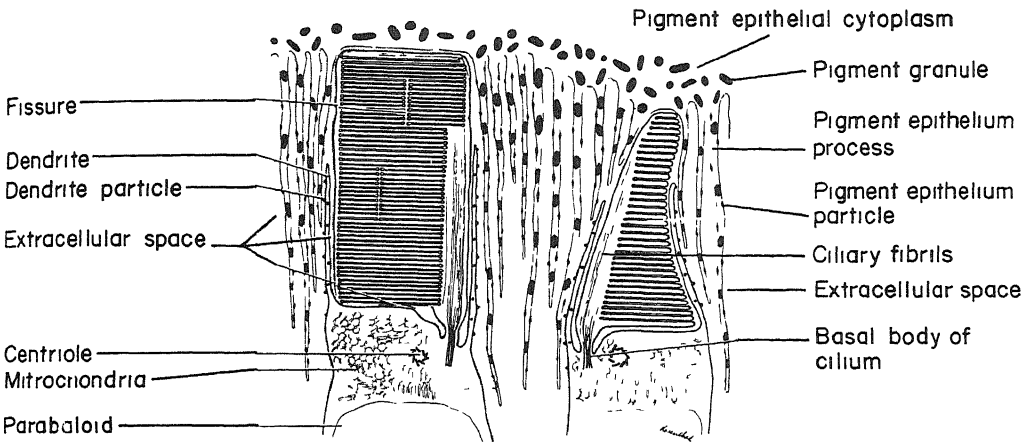


Fig 8 70 Reconstructions of the structures of rods and cones based on electronmicrographs of the receptors of *Necturus* Rods are on the left and cones are on the right All edges of the double membrane discs of the rods appear to have a differentiated rim structure In the cones the disc-like structures appear actually to be continuous in the foldings of a common plasma membrane and lack special limb structure fissures and lamellar micelles seen in the rods (de Robertis 1956)

these sacs as they are seen in the cones of the rhesus monkey The outer segments of foveal cones in the human are only 1 or 2 μ wide and 50 μ long The molecules of photosensitive substance are contained in the outer segments of the receptors The photochemistry of this system is treated below

THE EXTRAOCULAR MUSCLES

The eyes do not remain stationary in the head but move in a highly coordinated fashion Fixation can be maintained on a moving object only by relatively precise conjugate movements of the eyes With changes in the depth of an object viewed, the eyes must converge or diverge appropriately depending on whether the object is approaching or receding The immediate control of movement of each eye is effected by six extraocular muscles (Fig 8 72) Four rectus muscles control movements about two approximately perpendicular axes which are in turn nearly perpendicular to the line of sight with eyes directed straight ahead The lateral or external rectus is located on the temporal side of the eye, the medial or internal rectus is on the nasal side of the eye, with superior rectus at the top of the eye and the inferior rectus at the bottom of the eye Two additional muscles, the superior oblique and the inferior oblique, cause torsional rotation of the eye about the visual axis with some depression or elevation (Fig 8 73)

The Neural Connections of the Visual System

THE RETINA

The advent of electron microscopy has made possible a much better understanding of the nature of neural connections within the retina (Fig 8 74) Observations have been made on the human and on the monkey, cat, rat, rabbit, ground squirrel, and pigeon The basic patterns of cells are similar across species but the detailed nature of their connections differs Receptors make direct synaptic connection with bipolar cells and also with horizontal cells Horizontal cells in turn make synaptic connections with bipolar cells over a lateral extent such that they provide a mechanism for spatial interaction among different receptors A bipolar cell may make contact with a number of receptors directly in addition to receiving effects from more distant receptors via horizontal cells The inner terminations of bipolar cells are usually dual in nature and have been called *dyads* In man and the monkey, the dyad terminations from a single bipolar cell are usually on a ganglion cell and an amacrine cell There are considerable differences in the nature of these connections among species In the pigeon and the frog, there are relatively few direct terminations of bipolar cells on ganglion cells Information from bipolar cells is much more often mediated by

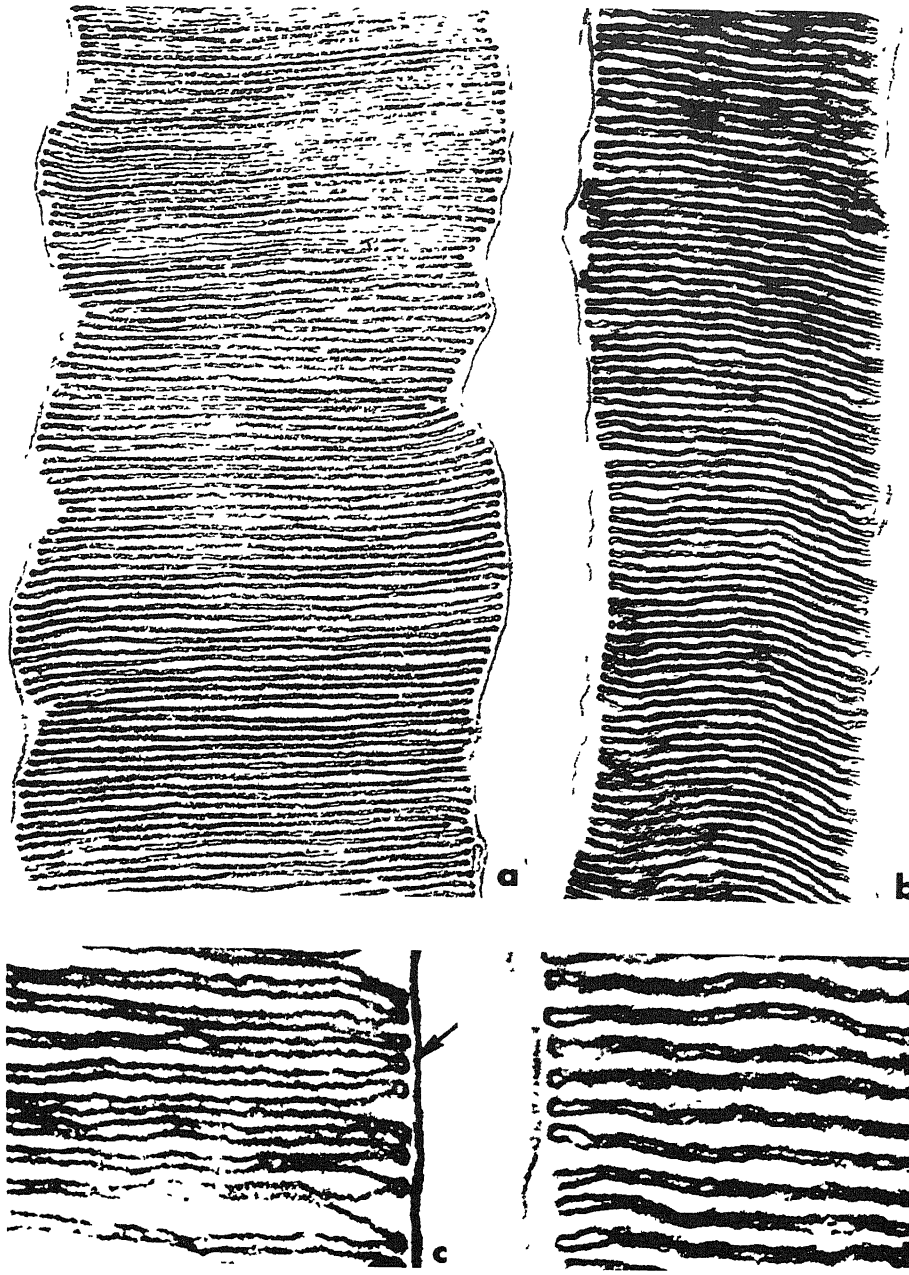


Fig 8 71 Electron microphotographs of longitudinal sections of the outer segments of a rat rod (*a*) and of a foveal cone of the rhesus monkey (*b*) The double membrane discs are similar in each of these structures showing no continuities with plasma membrane or with each other (amplification × 50 000) The outer segments of a rod (*c*) and a cone (*d*) of a monkey retina are shown at higher magnification (× 150 000) below The double membrane discs of the rod and cones differ only in that the rod membranes may be slightly thinner and their edges possess a characteristic rim structure as shown by the arrow (From Wald and Brown 1965)

amacrine cells This is true to a lesser extent in the ground squirrel and the rabbit It is still less so in the cat There may be little difference between the monkey and the human, with very many fewer bipolar to amacrine contacts, certainly in the region of the fovea, than is true for much lower animals Differences in synaptic connections among bipolar cells,

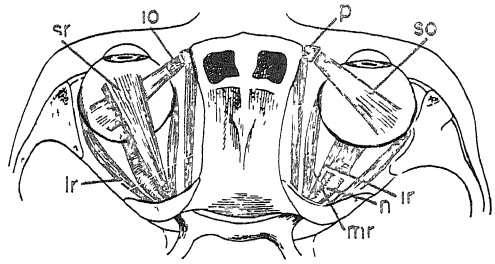


Fig 8 72 Oculomotor muscle of man as seen from above in a dissected head On the left a portion of the superior oblique has been cut away to reveal the inferior oblique on the right the superior rectus has been removed to permit a view of the inferior rectus io inferior oblique ir inferior rectus lr lateral rectus mr medial rectus n optic nerve p pulley for tendon of superior oblique so superior oblique sr superior rectus (From Walls 1942)

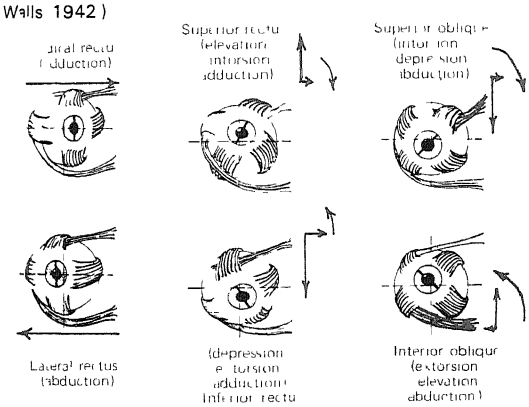


Fig 8 73 Action of the three pairs of antagonistic eye muscles (Adler 1965)

amacrine cells and ganglion cells which have been observed with the electron microscope may be correlated with functional differences in the neural activity of the retina as described in a later section

THE OPTIC NERVE AND OPTIC TRACT

The spatial organization of the retina is preserved in the pattern of orientation of fibers which leave the eye in the optic nerve (Fig 8 75) Fibers that represent signals originating in the temporal retina do not cross the midline at the optic chiasm but travel up into the optic tract on the same side of the brain Fibers originating in the nasal half of the retina, on the other hand, depart from those fibers which represent the temporal half, cross at the chiasm and travel centrally within the optic tract on the opposite side of the brain The implications of these connections have been tested by observing the effects of lesions made in the regions indicated in Figure 8 75 The major import of this anatomical design is that stimulation of receptors from objects within the right half of the visual field will stimulate the left half of each retina and will give rise to nerve signals which will be transmitted to the left hemisphere of the brain Conversely, stimulation from the left visual field ultimately is transmitted to the right hemisphere of the brain

The fibers within the optic tracts terminate in various places Some go to the pretectal region (Magoun and Ranson, 1935a and 1935b) and others to the superior colliculus (Davson, 1949) Most terminate in the lateral geniculate body

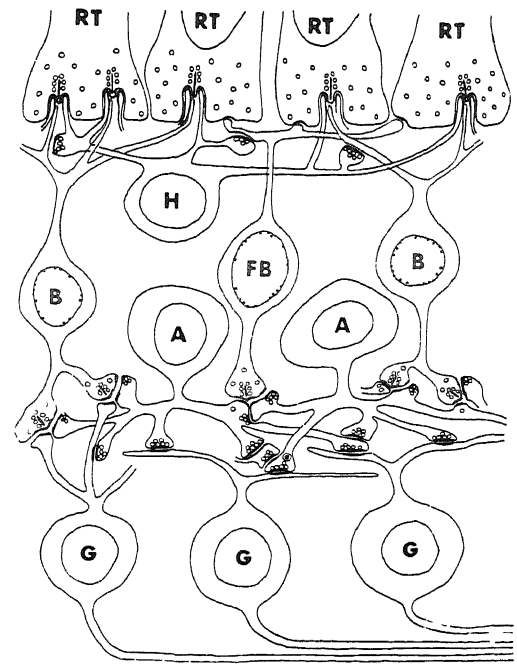


Fig 8 74 The synaptic contacts of vertebrate retinas The receptor terminals (RT) are penetrated by processes from bipolar (B) and horizontal (H) cells which terminate near the synaptic ribbons of the receptors Flat bipolar cells (FB) make superficial contacts on some receptor terminals Horizontal cells make synaptic contacts with bipolar dendrites as shown and with other horizontal cell processes as well Horizontal cells appear to mediate connections between receptors and bipolar cells over relatively great lateral separations on the retina In the region corresponding to the inner plexiform layer bipolar terminals are shown to contact either one ganglion cell (G) dendrite and one amacrine cell (A) process at ribbon synapses or two amacrine cell processes Amacrine cells make synapses with each other with ganglion cells and back onto bipolar terminals The relative number of direct contacts between bipolar cells and ganglion cells appears to increase the higher the level of the animal considered (Dowling 1970)

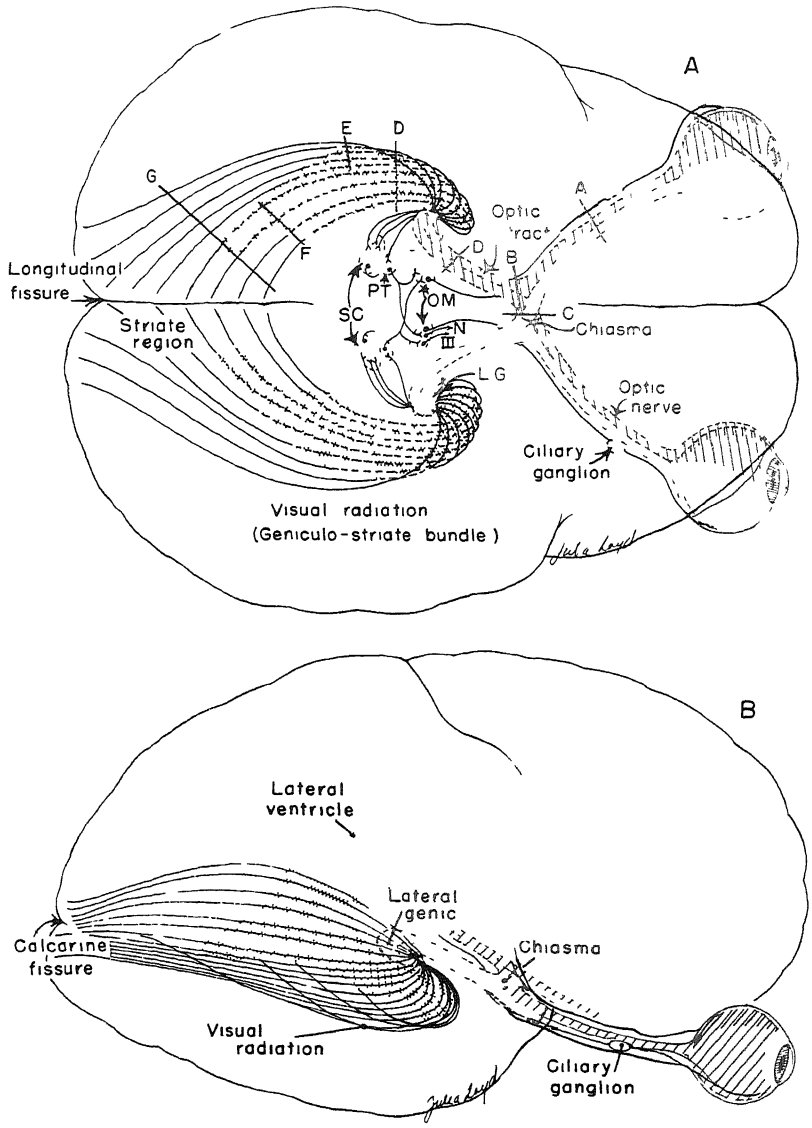


Fig 8 75 Diagram of the central visual pathways *A* plan view visual defects corresponding to lesions A B C D E F G are described in the text S C superior colliculus P T, pretectal region O M oculomotor nucleus L G lateral geniculate *B* elevation view (Brown 1965a, as modified from Homans and Fox and German)

THE LATERAL GENICULATE BODY

Cells of the dorsal nucleus of the lateral geniculate body are in six layers, the first, fourth, and sixth receiving fibers from the contralateral eye, and the second, third, and fifth receiving fibers from the ipsilateral eye (Fig 8 76) There is topological correspondence between points on the retina and areas of termination in the lateral geniculate body (Clark and Penman, 1934) Although there is some evidence for binocular fusion of neural activity at the lateral geniculate body for the cat (Hayhow, 1958, Sanderson et al , 1969), there seems to be little binocular interaction at this level in the monkey or, presumably, in man. Recent information from the laboratory of Peter Bishop suggests that there may be binocular interaction in cells of the lateral geniculate body If this is so, it is possible that it is mediated for at least one eye via cortical connections In



Fig 8 76 Section through the left lateral geniculate body of a rhesus monkey sacrificed 8 months after enucleation of the left eye (Brindley 1960 after Minkowski)

monkeys, neural fibers reaching the lateral geniculate body divide into several branches, each of which may terminate on a different neuron There is thus considerable divergence from retina to lateral geniculate The fact that there are approximately as many cells in the lateral geniculate body as there are fibers reaching it via the optic tract suggests that

there must also be convergence on single cells of the lateral geniculate body The idea that there is only a single input to each cell (Glees and Clark, 1941) is difficult to accept Various authors have suggested differential functions for the several layers of the lateral geniculate body (DeValois, et al , 1958, Clark, 1949) but the possible differentiation of their roles remains in doubt

CORTICAL CENTERS FOR VISION

The radiation fibers from the lateral geniculate body project to the striate cortex of Area 17 of the occipital lobe via the geniculostriate bundle This bundle, the visual radiation, travels forward and laterally then down around the ventricle of the temporal lobe before reflecting back and upward to the occipital cortex (Fig 8 75) That a topologic organization of the individual fibers with respect to retinal origin of their signals is maintained can be demonstrated by sectioning the optic radiation as illustrated in Figure 8 75 The effects of sections in various regions throughout the neural pathway are of interest As shown in the figure, a section of one optic nerve results in complete blindness of one eye (*A* in Fig 8 75) A transverse cut on one side of the optic chiasm (*B*) will result in blinding of the temporal half of the retina on the same side and possibly the nasal half or a portion of it for the retina on the opposite side If a medial section of the optic chiasm is made (*C*), both nasal retinas are blinded Section of the entire optic tract on one side (*D*) results in blinding of the half of the retina of each eye on the same side as that on which the section is made A section of the entire geniculostriate bundle has the same result If the upper bundle is cut on one side (*E*), there is blindness in the upper retinal quadrants in both eyes on the same side as the cut A cut in the lower portion of the bundle (*F*) produces blindness of the lower retinal quadrants on the same side

The striate area of the occipital cortex (Area 17) extends along the upper and lower lips of the calcarine fissure The way in which radiation fibers which represent various retinal regions terminate in the cortex is portrayed by the coding of the cortex and the retinal field as shown in Figure 8 77 There is a disproportionately large number of cortical cells which subserve the processing of information from the macular region of the retina It has been

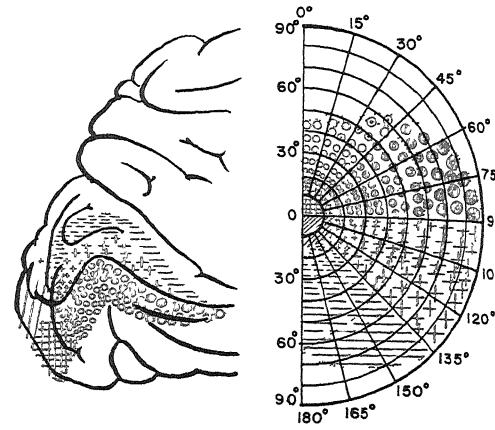


Fig 8 77 *A* projection of the retina on the visual cortex Various regions of the visual field (*B*) referred to regions in the striate area (*A*) as indicated by the coding The figure is drawn with calcarine fissure spread apart to reveal its inner walls (From Holmes 1945)

estimated that there are on the order of 100 cortical cells for every cone in the fovea of the eye of the monkey (Marshall and Talbot, 1942) The widespread representation of the central area of the retina renders vision for this part of the visual field relatively unaffected by even severe lesions of the occipital lobe Areas 18 and 19 of the occipital cortex receive projections from the striate area and have been assigned associative and motor functions (Chow and Hutt, 1953)

EFFERENT PATHWAYS

Efferent fiber pathways within the visual system have been well established anatomically in birds, but their presence in man is uncertain and their function, even when they are known to exist is not at all clear (Brindley, 1960) It has been shown recently that the relation between excitatory and inhibitory processes in the retina of the chick is disturbed by electrical stimulation of centrifugal fibers (Miles, 1970)

INTERMODALITY EFFECTS

A variety of stimuli from other sensory modalities such as hearing, smell, and the vestibular sense may influence the response of the visual system (Hernández-Peon, et al , 1957) Such influences may be mediated via the retic-

ular formation (Granit, 1955) There is a variety of evidence for involvement of regions of the nervous system not classically associated with the visual pathways in response to visual stimulation For example, units in the ventral medial hypothalamus of cats have been reported to show an increase in firing rate in response to visual stimulation by diffuse light or shadows (Campbell, et al , 1969) Several investigators have shown that unit activity in the amygdaloid nucleus is influenced by visual stimulation (Sawa and Delgado, 1963, O'Keefe and Bouma, 1969) A variety of responses to visual stimulation has been recorded from the inferotemporal cortex Recently, complex forms of response to visual stimulation have been recorded in single units of this area (Gross, et al , 1969) The posterior cingulate gyrus has been implicated with the visual system by the demonstration that effects on cortical dendritic branching in this region may occur as a result of dark rearing (Coleman and Riesen, 1968) The red nucleus is implicated in visual processes by reason of the fact that its destruction may cause a deterioration of the ability to make visual size discriminations (McNew, 1968) As mentioned above, inputs from a number of other sensory modalities may influence visual responses The simple expedient of tilting the body appears to change the orientation of receptive fields for visual stimuli in the occipital cortex (Horn and Hill, 1969)

There is a variety of evidence that regions of the nervous system which have classically been assigned an exclusive role in connection with some sense modality may, in fact, be involved in a variety of sensory functions Monkeys which had been peripherally blinded earlier were able to perform a locomotor maze task in which they were required to move through the maze After lesions were made in the striate cortex their performance was impaired Their performance in a stylus maze task in which total body movements were not required showed no impairment (Ohrback, 1959) The implication is that various regions of the cortex, including those classically assigned a visual function, may play an important role in the organization and control of behavior in the absence of any stimulation via the classically associated sensory modality Knowledge of the functional anatomy of the nervous system can be characterized only as primitive at this time

PHYSIOLOGY OF VISION

Photochemistry

The initial transformation of radiant energy in the eye which results ultimately in vision involves the absorption of light quanta by molecules of photosensitive substance It has been known for years that the light sensitive material for rods in many organisms including the human is rhodopsin (Wald and Brown, 1958) The relative spectral response of the eye when dark adapted or under conditions of low illumination for regions of the retina which abound in rods agrees well with the absorption spectrum of this pigment as shown in Figure 8 78 (cf , V'_λ, Fig 8 57) The photochemistry of bleaching and regeneration of rhodopsin in the visual process has been worked out in great detail by Wald and his colleagues (Wald and Brown, 1965)

The visual pigments involved under conditions of higher light illumination, those pigments which reside in the cones, are not understood nearly so well as is rhodopsin It had long been assumed that there were perhaps

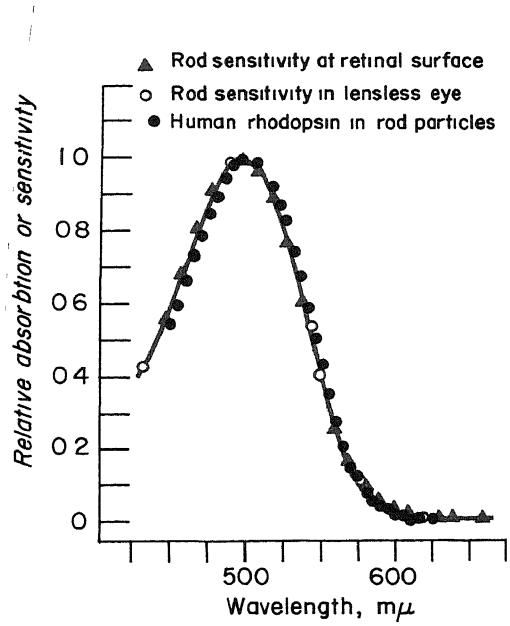


Fig 8 78 Absorption spectrum of human rhodopsin measured in a suspension of rod outer segments compared with the spectral sensitivity of human rod vision The latter data have been corrected for ocular transmission or are based on measurements of the spectral sensitivity in aphakic (lensless) eyes (From Wald and Brown 1958)

three pigments, based on the trichomacy of vision (discussed in a later section) Only recently, however, have we had fairly direct evidence as to the existence of three different classes of pigment which probably reside in each of three different classes of cone receptors The spectral response function of the light adapted eye, measured in the fovea, provides a basis for the specification of the units of light as discussed earlier It also is believed to represent some form of composite absorption spectrum for the cone pigments

SPECTRAL ABSORPTION FUNCTIONS OF VISUAL PIGMENTS

Several different approaches to the determination of the nature of pigment absorption characteristics have been taken in addition to extraction and in vitro measurement First, it is possible to measure the spectral response function V_λ, as illustrated in Figure 8 57 This is the composite function for all cone pigments If the eye is then exposed to a steady background illumination of intense blue light which tends to bleach out the pigments with peak absorption in the middle and short wavelength regions, the resulting spectral response function may be assumed to represent the absorption spectrum of a long wavelength pigment The absorption spectrum of a pigment which peaks in the middle wavelength may be revealed by measurements against a background of mixed long and short wavelength, or a purple light The nature of the spectral absorption function of a pigment which absorbs maximally in the short wavelength region may be inferred from measurements made with the eye adapted to a highly intense yellow light which bleaches out pigments of both long and middle wavelength absorption spectra The results of such measurements are discussed in detail in the treatment of color vision

Measurements of this kind show an approximate agreement with other measurements based on the computation of difference spectra after selective bleaching (Brown and Wald, 1963) They are also in agreement with results which have been obtained in the living human eye by an objective measurement of the light reflected back out of the eye for various wavelengths of illumination following selective bleaching When appropriate account is taken of the fact that such light has made two passes through the retina, it is possible to compute the absorption spectrum for the unbleached eye, and, after bleaching, difference spectra which may represent individual pigments Although this technique has the advantage of providing objective records on the living human eye, the interpretation of results involves a number of complications (Rushton, 1964)

Another interesting and extremely difficult technique involves the measurement of spectral absorption characteristics of pigments within individual cells The technique is known as microspectral densitometry Its difficulties are such that only a few investigators have undertaken it and data are available on only a very few individual cones from the human retina (Marks et al, 1964, Brown and Wald, 1964) Those data which are available tend to agree with data based on selective adaptation or selective bleaching of the human retina One set of curves which provides a representative summary of the relations among the absorption spectra of what are believed to be three retinal cone pigments and the composite cone sensitivity function is presented in Figure 8 79 The data points in the figure are based on measurements at the cornea When selective spectral absorption of various ocular media is taken into account, the relative shapes of the cone functions are changed slightly This is shown on the right half (*B*) of Figure 8 79 The primary difference is an elevation in the relative amplitude of the short wavelength function This results largely from the blue absorbing properties of the yellow macular pigment The peak values for the functions presented are at 430, 540, and 575 nm The hues which correspond with these wavelengths are a blue, a green, and a yellow Thus, the long wavelength "red" pigment really has its maximum absorption in the yellow region of the spectrum, although it extends well into the longer wavelength red region

PHOTOCHEMICAL CYCLES

The photochemical cycles for a large variety of visual pigments in a number of organisms appear to be remarkably similar (Wald, et al, 1963) Each pigment is identified with its own specific type of protein, an opsin which is united with the eleven-*cis* isomer of retinaldehyde Retinaldehyde is the chromophore of the complex molecule and it is on this portion of

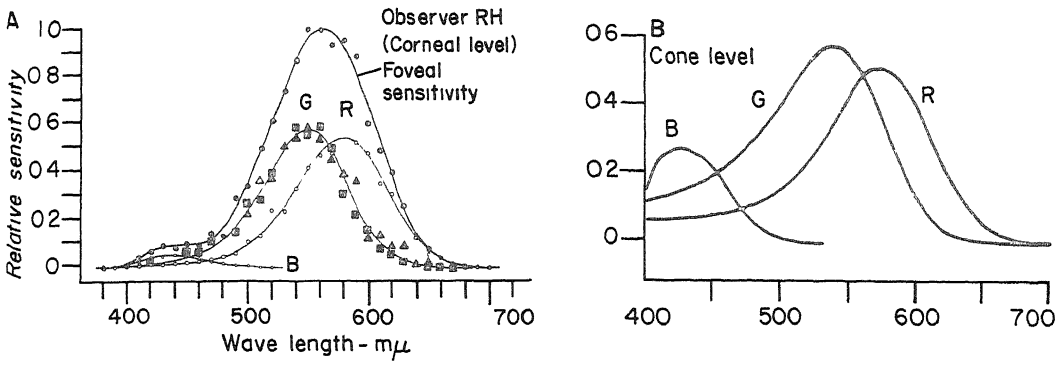
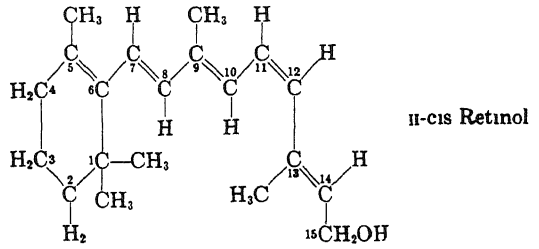
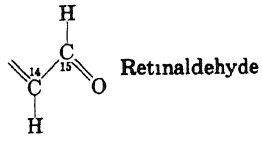


Fig 8 79 *A* individual color receptor mechanisms and total foveal sensitivity of observer RH The relative amplitudes of the sensitivities of the individual receptor mechanisms are based on measurements at the corneal surface *B* relative amplitudes as they would be found at the cones In the average observer the blue component would be only about one-third as high as the amplitude found in RH (From Wald and Brown 1965)





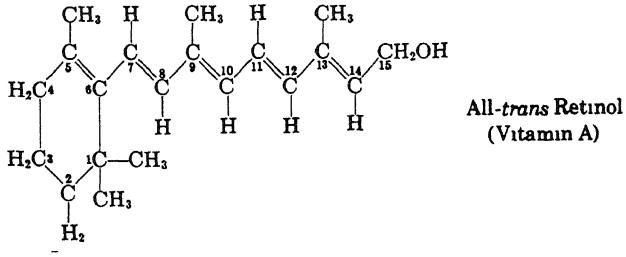


Fig 8 80 Molecular structures of 11-*cis* Retinol retinaldehyde and all-*trans* Retinol (vitamin A)

the molecule that the light acts The nature of the action is solely to transform retinaldehyde (illustrated in Fig 8.80) from the eleven-*cis* to all-*trans* configuration (Hubbard and Kropf, 1958) The all *trans* retinaldehyde is initially still bound to the unmodified opsin of the photopigment This molecule is extremely unstable at ordinary temperatures, however, and in darkness is transformed rapidly over a series of intermediate products In the rhodopsin cycle, the first product of bleaching is known as prelumirhodopsin, which is transformed rapidly into lumirhodopsin The next stage is metarhodopsin I, then metarhodopsin II, and finally, as a result of hydrolysis, a mixture of opsin and free all-*trans* retinaldehyde. The last stage is so slow that it cannot play a direct role in visual excitation. Therefore, excitation

must depend upon one of the intermediate stages In the presence of an enzyme, alcohol-dehydrogenase, the all-*trans* retinaldehyde is reduced to retinol (vitamin A) Regeneration of the original retinal pigment requires a re-oxidation of retinol to retinaldehyde along with isomerization to the eleven-*cis* configuration A schematic illustration of the various stages of bleaching is presented in Figure 8 81 for rhodopsin The stages are probably similar for other pigments except that the unique opsins which characterize the pigments replace the opsin portion of the rhodopsin molecule Another form of retinaldehyde which differs from that found in mammals has been found in fish The difference consists in an extra double bond in an ionone ring Combination of the various opsins which have been found in naturally occurring photopigments with both of the types of retinaldehyde has resulted in the production of a variety of photopigments including some which have not been found in nature

The spectral absorption characteristics of photopigments when plotted on a wave number axis have all been shown by Dartnall (1957) to be of approximately the same form That is, if they are shifted along the wave number axis, they may be superposed The absorption spectra of these same pigments plotted on a wavelength axis are broader, the longer the wavelength at which they absorb maximally The examination of the way in which pigments combine and a variety of calculations may be simplified if absorption spectra are expressed with respect to the wave number axis

The bleaching and regeneration cycle does not invariably go through the stages outlined above A portion of certain unstable intermediate products is isomerized back into photosensitive forms by the direct action of light (Dowling and Hubbard, 1963) This rapid regeneration of a photosensitive substance from the unstable photoproduct may account for the fact that it is never possible to bleach more than half of the rhodopsin in the retina for relatively short duration exposures of a bleaching light, no matter how intense it may be (Hagins, 1956)

ABSOLUTE RETINAL SENSITIVITY

It has been estimated that the number of pigment molecules in an individual rod may be 70,000,000 or more (Pirenne, 1962) for each of the 120,000,000 rods in the human retina Each of these molecules may catch a quantum of energy It is of importance for our understanding of how the visual system works to examine the question of how much energy is necessary for the excitation of a conscious response to retinal illumination A number of investigators have studied this problem over the years, but not until around 1940 was an experiment performed which carefully considered all of the variables which must be controlled for the measurement of a truly minimum energy requirement (Hecht et al , 1942)

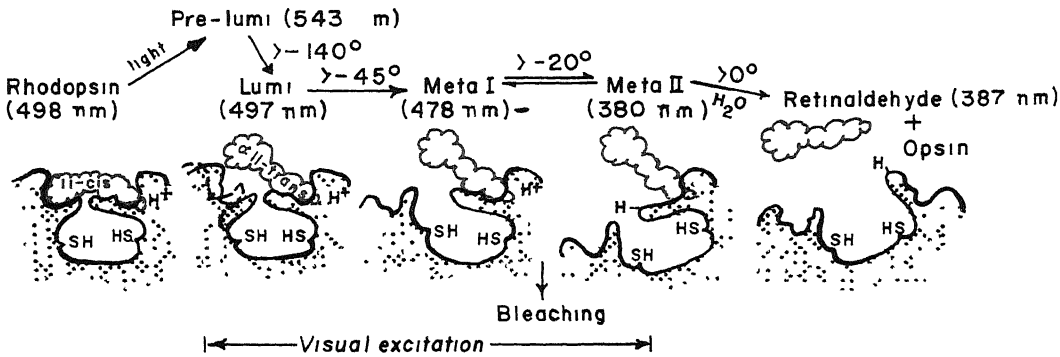


Fig 8 81 Stages in the bleaching of rhodopsin The chromophore of rhodopsin 11-*cis* retinaldehyde fits closely a section of the opsin structure The only action of light is to isomerize retinaldehyde from the 11-*cis* to the all-*trans* configuration (prelumirhodopsin) Then the structure of opsin opens progressively (lumi- and the metarhodopsins) ending in the hydrolysis of retinaldehyde from opsin Bleaching occurs in going from metarhodopsin I to II and visual excitation must have occurred by this stage The opening of opsin exposes new chemical groups including two -SH groups and one H+-binding group The absorption maxima shown are for prelumirhodopsin at −190° C lumirhodopsin at −65° C, and the other pigments at room temperature (Wald and Brown 1965)

The eyes of observers were dark adapted for at least thirty minutes in order to ensure maximum sensitivity of the rod receptors A test light wavelength of 510 nm was selected because this wavelength corresponds approximately with the wavelength of maximum absorption of the photopigment of the rods, rhodopsin The test flash illuminated a region of the retina located at a distance of 20° temporally along the horizontal meridian of the retina In this region the density of rods is at a maximum and there is no danger of complication by the blind spot (Fig 8 62)

Retina Image Size. The size of the image of the test flash on the retina was restricted to 10 min of arc in diameter It has been demonstrated that with increase in the area of a test patch on the retina, retinal illumination, expressed in lumens per unit area, may be reduced Within limits, the eye is capable of summation of light energy over its surface Thus, if a given total energy is required for stimulation of the eye, a lower number of lumens per unit area may be required if a larger area is illuminated (Bartlett, 1965a) This summation capability begins to break down if the area becomes too large, however There is a limit to the extent to which light may be dispersed over the retina without a reduction in the summation of its effectiveness The area selected by Hecht, Shlaer, and Pirenne (1942) was well within the limit for complete retinal summation

Test Flash Duration The flash duration which they employed was 0 001 sec Just as the eye is capable of summating energy over area, so it is also capable of summating energy in time Within limits, the longer the duration of a test flash, the lower will be the threshold level expressed in lumens The lumen is a unit of luminous energy per unit time or luminous flux Thus, the longer the eye is exposed to a given luminous flux, the greater will be the total energy of the exposure If threshold depends upon a fixed amount of energy, then the longer the flash duration, the lower will be the luminous flux necessary to achieve the constant energy required for threshold (Fig 8 82) This is true as long as a critical duration, beyond which temporal summation is no longer possible, is not exceeded (Graham, 1965a) Critical duration in the dark-adapted eye is approximately 0 1 sec The flash duration of 0 001 sec was therefore well within the critical duration

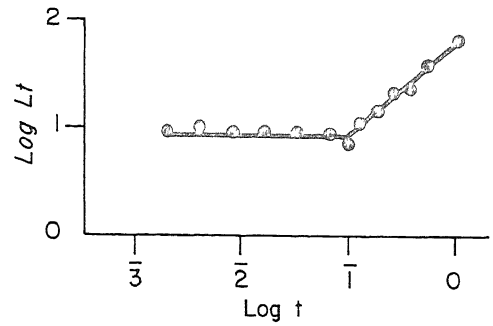


Fig 8 82 Logarithm of the product of luminance and flash duration as a function of the logarithm of flash duration for a small test light in the fovea Time is in seconds The product of luminance and time or energy remains constant up to a flash duration of 0 1 sec and then increases in direct proportion to duration (Graham 1965a)

Minimum Energy at the Retina. The apparatus used for the experiment was carefully calibrated in terms of radiant energy per unit time per unit area delivered to the corneal surface of the eye of an observer Precautions were taken to insure that the cross-sectional area of the beam entering the eye was known and was in no way restricted or altered by pupillary contraction Threshold measurements revealed that the test flash could be detected at significantly better than a chance level when between 54 and 147 quanta were supplied at the corneal surface By taking into account the absorption and reflection characteristics of the ocular media and the percentage of energy reaching the retina which is actually absorbed by the photopigment, it was estimated that the number of quanta necessary for excitation of the visual process in a human observer lies somewhere between 5 and 14 The best available current estimate of the number of rods in the retinal area which was stimulated is 350 The 5 to 14 quanta may be expected to be distributed randomly over such an area and, consequently, there is a negligible probability that more than a single quantum would be absorbed by a single rod at threshold

This is a highly significant finding It means that the action of a single quantum on one out of more than 70,000,000 molecules within a given receptor gives rise to events in that receptor which influence succeeding elements in the nervous system and ultimately arouse a conscious response The fact that not one but at least a small number of rods must be excited in the same short interval of time is also

significant Some molecular transformations may be expected to occur in the absence of light The requirement that a finite number, perhaps five, occurs within a short time interval provides the organism with a basis for the discrimination of true signals against the background noise level of the retina

Sources of Variability It is appropriate to mention one additional implication of this experiment It was observed that the variability of response of the subjects was only very slightly greater than the quantum variability to be expected in repeated presentations of the stimulus This implies that under very carefully controlled conditions, the physical variability of the stimulus may represent a major contribution to the total variability of the stimulus-response system

Physiology

ELECTRICAL ACTIVITY

The Electroretinogram. The recording of gross electrical activity of the retina in response to light stimulation has been possible for a number of years (Riggs, 1958, Gouras, 1970) If a voltage is measured between the front surface of the eye in the region of the cornea and a reference electrode located on the forehead or some region of the face, stimulation of the eye by light results in a complex wave form of response which occurs after a relatively short latency (Fig 8 83) A contact lens is used to position an electrode over the cornea Certain components of the wave form are associated with scotopic or rod processes and other components are associated with photopic or cone processes The wave form changes depending on the condition of adaptation of the eye when the stimulus is presented, whether a single stimulus or a train of stimuli is presented, and such variables as the luminance and wavelength distribution of the stimulus Although the components of the electroretinogram (ERG) may represent rod and cone processes, this does not necessarily imply that they reflect voltage changes within the receptors themselves The latency of response of some of the components is much too long for that and it is probable that the major components of the ERG, as it has been typically recorded, reflect events which are triggered by the stimulation of receptors but which occur in subsequent elements of the visual pathway within the retina

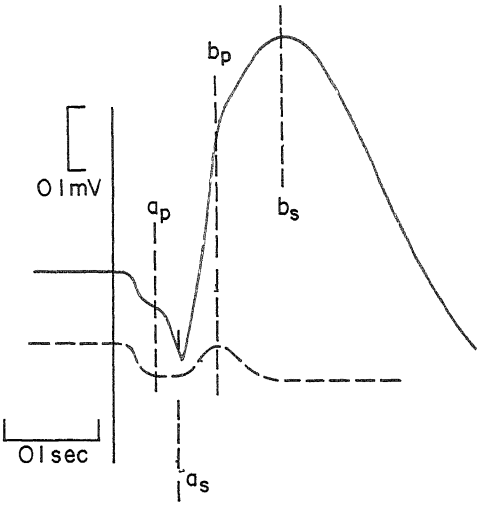


Fig 8 83 The electroretinogram in the dark-adapted human eye (solid line) as the resultant of photopic (dashed line) and scotopic (dotted line) components The a-wave is composed of the photopic (a_p) and scotopic (a_s) components and the b-wave is composed of the photopic (b_p) and scotopic (b_s) components (Armington et al 1952)

The Early and Late Receptor Potentials In recent years (Brown, 1968) it has been demonstrated that electrical activity of the retina, which almost certainly has its origin in the receptors themselves, may be recorded In animals, these components may be seen when an electrode is inserted into the eye and into the retinal layers an appropriate depth The voltage change which occurs in response to light stimulation is measured with respect to a second reference electrode, the tip of which is located in the vitreous humor of the eye Under these circumstances two components are revealed, one of which is of almost zero latency and the second of which occurs approximately 1 msec after the stimulus These have been called the early and late receptor potentials (ERP and LRP), respectively (Fig 8 84) The first is associated with the physical events which accompany the absorption of quanta by the photopigment The significance of the late receptor potential is unknown, but presumably it is associated with the events which result from quantal absorption and which precede or cause excitation of the neural elements in the system

The early and late receptor potentials of the retinal response can be observed in human subjects if an appropriate technique is em-

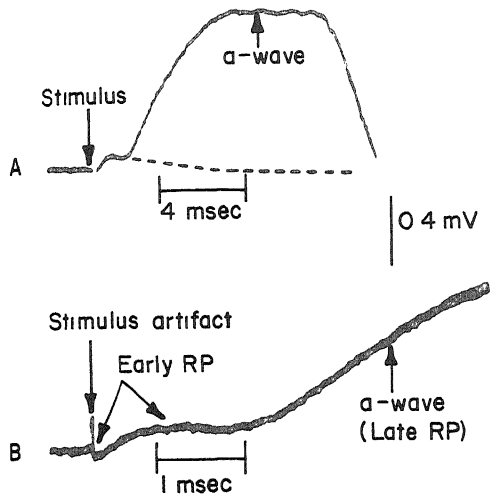


Fig 8 84 The early and late receptor potentials of the cynomolgus monkey The top record *A* is from the peripheral retina and shows the early response between the occurrence of the stimulus flash and the rising phase of the a-wave The course of decay of the early response which can be isolated by cessation of artificial respiration is shown by the dashed line The bottom record *B* is from the fovea where the response is somewhat larger in amplitude The stimulus artifact is only positive and thus the early receptor potential is seen to be biphasic The late receptor potential corresponds to the rising phase of the a-wave (Brown 1968)

ployed Obviously, it is not possible to pierce the eyeball with a recording electrode A conventional method of recording the electroretinogram may be combined with a computer averaging technique for the purpose, however An electrode is mounted in a contact lens such that it can be located over the corneal surface of the eye A light flash is presented repeatedly and the response to each successive flash is accumulated in an appropriate computer system This process is synchronized with the time of presentation of the flash Those electrical events which invariably occur at a certain time following the initiation of a flash tend to summate as the number of flashes is increased Random events which are not directly correlated with the flash tend, on the other hand, to cancel each other out It is thus possible, after a sufficient number of presentations, to visualize components of the electrical response which could never be seen by recording only a single response

Changes in the early receptor potential are approximately proportional to changes in the physical energy of stimulation On the other hand, changes in the late receptor potential are, over a substantial range, proportional to the logarithm of the stimulus intensity Stimulus intensities may be reached beyond which the response no longer increases This is to be expected when levels of illumination are employed which result in substantial bleaching of the photopigment The interesting point is that the late receptor potential shows a relation to the stimulus which is similar to the relation between response and stimulus of a human observer in psychophysical experiments The effectiveness of an increase in stimulation is more nearly directly related to the ratio of the increase than the absolute amount of the increase There is also a breakdown of this relation as higher stimulus levels are achieved The conclusion which may be drawn is that significant aspects of the transformation which occurs between peripheral stimulation and a complex response to sensory inputs occur at the level of the receptors themselves Additional implications of this finding are considered in connection with adaptation of the eye

Recently, it has been demonstrated that the early receptor potential is dominated by the cones in human vision (Goldstein and Berson, 1969, 1970) Changes in the early receptor potential which are associated with inherited retinitis pigmentosa have also been found (Berson and Goldstein, 1970)

RETINAL INTERACTION EFFECTS

There are many lateral interconnections among the pathways from individual receptors to the layer of ganglion cells and subsequently to higher visual centers of the brain (Fig 8 72) If a spot of light stimulates one retinal region, effects of such stimulation may occur in surrounding regions as a result of the activity of horizontal or amacrine cells It would be of very great interest to be able to record from each of the classes of cells in the retina during stimulation in order to determine the way in which their responses compare and what sort of interaction mechanisms exist Unfortunately, this is a very difficult task Intracellular recording with electrodes of very small diameter (less than 1 μ) has been accomplished in each of the various types of cell in the retina of *Necturus*, the mud puppy The job is easier in *Necturus* by reason of the relatively large size of the individual cells (Fig 8 85) A black bar in this figure represents illumination of one receptor The response of the receptor is a

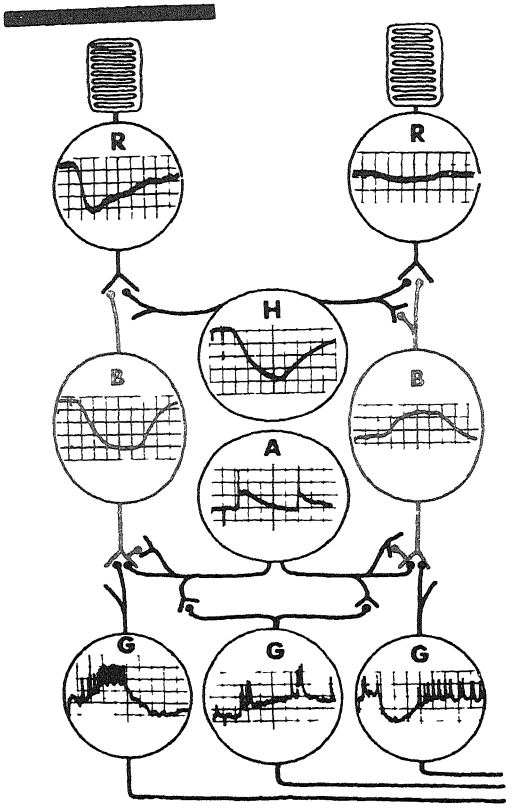


Fig 8 85 Intracellular responses of various retinal elements to stimulation of an isolated retinal region R receptors H horizontal cell B bipolar cells A amacrine cell G ganglion cells The stimulation is indicated by a black bar over the left receptor The responses of the receptor stimulated and the bipolar and horizontal cells in direct contact with it are shown to be hyperpolarization The action of the horizontal cell on the bipolar cell which it contacts is apparently to cause depolarization Spikelike discharges are first seen at the level of the amacrine cells but the spikes are single transients which occur only at onset and termination of illumination Typical neural spike discharge is first seen at the level of the ganglion cells (Dowling, 1970)

graded potential and varies in level with respect to the amplitude of stimulation Although the initial response may be somewhat higher than the final sustained response, the response does not decay completely as long as the stimulus is left on Perhaps surprisingly, the nature of the response is one of hyperpolarization Instead of a reduction in the level of the membrane potential, this is increased Hyperpolarization appears to be typical of the receptors themselves, the horizontal cells and the bipolar cells Although studies of neural tissue would not have led one to predict this outcome, one must infer that hyperpolarization of certain elements within the nervous system may somehow be associated with excitation of succeeding elements with which they are in contact

Lateral Inhibition. A receptor displaced laterally from that which directly receives stimulation may show little or no response Such response as is observed could be attributed to the effects of scattered light within the eye The bipolar cells connected to the adjacent receptors do show a response, however, and it is of opposite polarity to that of the bipolar cells which are directly connected with the stimulated receptor This lateral projection onto bipolar cells would thus appear to be inhibitory in nature The time course of the effect is similar to that of the effect measurable in the receptors themselves and in the horizontal cells and hence the pathway for such lateral inhibition would seem to be via the horizontal cells Amacrine cells provide a horizontal or lateral linking pathway, but they respond with a transient It is thus difficult to understand how they might mediate the graded depolarization observed in adjacent bipolar cells

Some bipolar cells have been found in the mud puppy for which the graded potential is in the direction of depolarization The response of either type of bipolar cell to a small spot of light stimulation is reduced or antagonized when additional illumination is added around the region of the central spot There thus appears to be a kind of antagonistic interaction between cells which serve adjacent regions of the retina

Other Interactions. The amacrine cells do show a transient response more nearly like that of a typical nerve cell, but they appear to respond with a transient at onset and offset of illumination, rather than continuously The ganglion cells are the first cells which show repetitive spike discharge, the frequency of which is related to the amplitude of stimulation Amacrine cells may have some role in adaptation and almost certainly are involved in the complex processing of visual information in the retina of lower vertebrates (Lettvin et al, 1959)

Insofar as Dowling and his colleagues have been able to make recordings from individual cells of the retina in other vertebrate species higher than *Necturus*, they have not found contradictory results Nonetheless, it would be

premature to generalize from their finding to the mammalian retina or to the primate retina without some qualification As discussed earlier, the nature of interconnections among the various cell types of the retina differs significantly even among mammals

RECEPTIVE FIELD ORGANIZATION OF THE RETINA

Long before the technical problems associated with intracellular recording in the retina had been overcome, it was possible to record the activity of individual ganglion cells with extracellular electrodes (Kuffler, 1952) In a typical procedure, the electrode is inserted into the retina until activity which represents the discharge of a single cell is detected The activity may be spontaneous, or some search technique involving a flashing light or a moving pattern may be employed Once a cell response is identified, the electrode is maintained stationary and an exploration of the way in which response of the cell changes through stimulation of the retina in various regions may be undertaken When explorations are made with a small test spot, it is found that the retinal region which must be stimulated to induce any activity or change in activity at the cell is roughly circular It is divided between a central zone and a surrounding zone which are antagonistic in their effects Stimulation of the central zone may cause either excitation or inhibition For cells which are excited by stimulation of their central region, the stimulation of the surround causes inhibition When stimulation of the central region is inhibiting, stimulation of the surrounding region will be exciting (Fig 8 86)

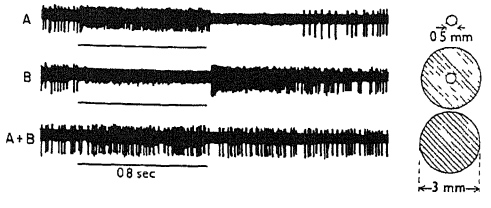


Fig 8 86 The spike discharge response of a ganglion cell of the retina of a cat to various forms of light stimulation An appropriately centered small spot of stimulation gives rise to a high frequency discharge Stimulation of an annular region around the central spot results in inhibition Stimulation of the entire region both center and surround results in a level of discharge slightly less than that evoked by the center spot alone (Brown and Wiesel 1959)

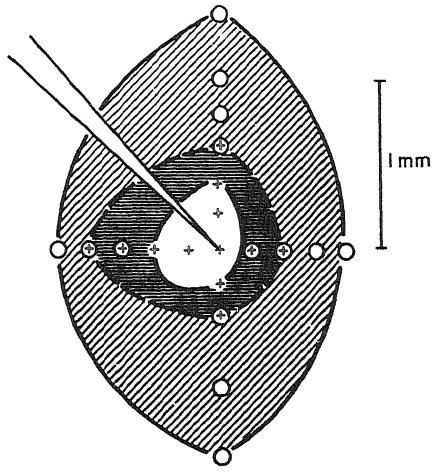


Fig 8 87 Diagram of a receptive field in the retina of the cat The contours are based on the response of a microelectrode which is presumed to be in contact with a single ganglion cell When an exploring light spot illuminates the region labeled + *on* discharges occur when regions labeled O *off* discharges occur There is an intermediate region marked ⊕ in which both *on* and *off* discharges are recorded (From Kuffler 1953)

This kind of receptive field organization is found in the cat (Fig 8 87), in the monkey, and presumably is the type of receptive field organization which exists in man This organization is also found in other animals such as the rabbit, the squirrel, the frog and the pigeon, but other more complex organizations are also present (Lettvin, et al 1959) These more complex organizations are apparently associated with a greater significance of the role of the amacrine cell in retinal interaction

There is some evidence that the organization of the receptive field may change when the eye is dark adapted (Studies of receptive field organization are typically performed in a light-adapted preparation) Barlow and co-workers (1957) have reported that the antagonistic surrounding element of the receptive field drops out when the eye becomes dark adapted Their experiment has attracted considerable interest in connection with our efforts to understand the nature of the dark adaptation process To some extent at least, the increase in sensitivity of the dark-adapted eye must be attributed to an increased spatial summation capability Lateral inhibitory effects are not entirely compatible with spatial summation and the notion of a reorganization of the receptive field during dark adaptation which might favor summation is therefore attractive (Fig 8 88)

Just as the absence of lateral inhibition effects may render the eye more capable of spatial summation and hence increase its sensitivity at very low light levels, lateral inhibition at high light levels may be useful as a basis for enhancing the spatial resolution capacity of the eye Excitatory receptive field centers on the lighter side of a boundary between a light portion and dark portion of the retinal image such that the inhibitory surround is partially within a darker region will have a lesser total amount of light on the inhibitory surround than is true for other receptive fields of similar spatial size and organization which are all in the region of higher illumination In consequence, the ganglion cell associated with the receptive field center near the boundary with the darker region may be excited to a higher level of activity The effect may be to enhance the difference in excitation selectively in the region of boundaries (Fig 8 89)

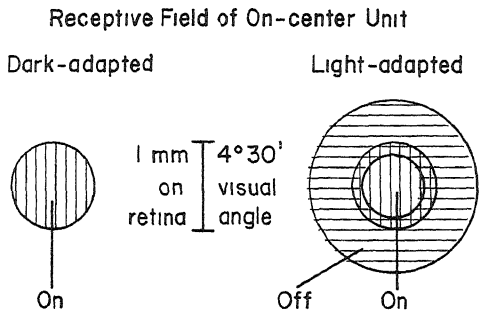


Fig 8 88 Suggested organization of the receptive field of an *on*-center unit in the retina of the dark- and light-adapted cat *On* or *off* responses are obtained when the stimulus falls in the correspondingly labeled regions of the field Locations of the limits of these regions are based on inferences from area threshold curves (Barlow Fitzhugh and Kuffler 1957)

Attempts have been made to devise precise quantitative models which will be descriptive of the role of inhibitory processes in the retina (Reichardt and MacGinitie, 1962) When such models adhere strictly to known principles of neuroanatomy and neurophysiology, involve constants which have values which are reasonable in biophysical or biochemical terms, and at the same time provide a good description of experimental results and possibly afford some prediction of what one may expect in experiments yet undone, then they may be considered as theoretical interpretations of what may actually be occurring in the nervous system Unfortunately, most such models fit only the data upon which they are based, require arbitrary constants to do so, and are of limited predictive value

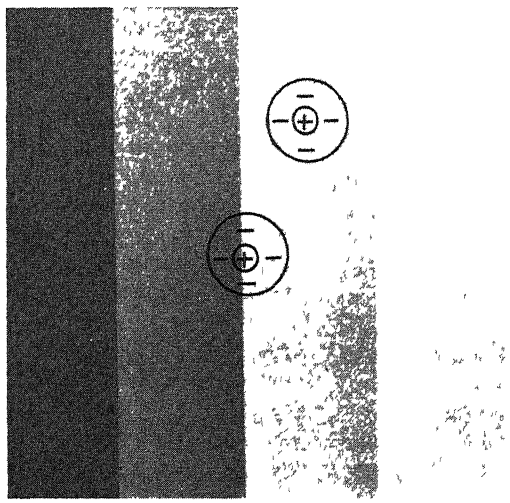


Fig 8 89 Border contrasts in a pattern of bands which increase in lightness from left to right The reflectance of each individual band is constant although bands appear lighter in the region of the left-hand boundary This may result from the lesser amount of stimulation of surrounding inhibitory regions of receptive fields in the vicinity of the boundary In the schematic illustration of receptive fields on the figure the ganglion cell stimulated by the field adjacent to the boundary will presumably fire at a rate higher than that stimulated by the field located near the center of the band

DIFFERENTIAL RETINAL RESPONSES TO STIMULUS WAVELENGTH

In animals with color vision, there must be some differentiation of responses to stimulating lights having different wavelength distributions Such differentiation must occur at the most peripheral point in the receptor system As noted above, cones are equipped with pigments having different absorption spectra and hence may respond differentially when the wavelength of the stimulus is changed It may be assumed that a different pattern of subsequent retinal events will follow when the light stimulus is changed from red to green, for example Unfortunately, most of the work which has been done on the differential responses of the retina to different wavelength distributions of the test stimulus has been done in the fish The color vision of the fish, although it must have arisen via a quite different evolutionary path from that of the human, is very similar to the color vision of man As illustrated in the first figure of Chapter 1, studies of the response of the fish retina to different wavelength distributions of the stimulus have shown that graded potentials may be obtained with a different polarity depending upon the spectral character of the stimulus (Svaetichen and MacNichol, 1958) Such graded potentials may have arisen in bipolar cells or horizontal cells It seems doubtful that the primary receptor itself could respond directly with either hyperpolarization or depolarization although this kind of response is not a physical impossibility. There are photoconductive materials which respond differentially to light of different wavelength distributions It seems probable that bipolar cells, such as those illustrated in Figure 8 85, may respond in one direction when light of an appropriate wavelength illuminates the receptor element to which they are directly connected and in a different direction when a nearby receptor having a different photopigment is stimulated by light of a different wavelength with a resulting activity in some horizontal cell which results in a response of the bipolar cell in the opposite direction

One study of the ganglion cells of the monkey indicates that there are indeed differential responses of these cells to changes in the distribution of the wavelength of the stimulus (Gouras, 1968) This work is preliminary, however, and does not provide a basis for our understanding of the detailed nature of the retinal connections which are involved in color discrimination

THE LATERAL GENICULATE BODY

The results of microelectrode studies of receptive fields for single cells of the lateral geniculate body of the cat (Hubel and Wiesel, 1961) are very similar to results obtained by electrode recording from the ganglion cell layer of the retina or the optic nerve Receptive fields show a concentric organization with antagonistic center and surround characteristics The variability of response at the level of the lateral geniculate body is considerably greater than is the case at the retina, however Activity within the reticular formation probably influences activity of the lateral geniculate body (Hubel, 1960, Arden and Soderberg, 1961) In addition, there is some evidence for cortical influence on activity of the lateral geniculate body (Widén and Ajmone-Marson, 1961) and anatomical bases for such influence have been confirmed Although binocular interaction within the lateral geniculate body has been reported in the cat (Bishop and Davis, 1953), there is no substantial evidence for physiological interaction of inputs from the two eyes at the level of the lateral geniculate body in primates In spite of the large number of experiments which have been performed on the neural activity of the lateral geniculate body, possible differential functions of the different layers remain obscure There is both convergence and divergence of connections from cells in the retina to cells in the lateral geniculate body and one result would appear to be an increase in spatial contrast at the lateral geniculate body (Hubel and Wiesel, 1961)

Color Coding The most extensive electrophysiological studies of the color vision process in primates have been done on the lateral geniculate body (DeValois, 1965) Individual cells have been found which respond by increased activity to all wavelengths of light and which have a spectral response characteristic similar to that of the visual system as a whole (Fig. 8 57) Other cells, called color-opponent cells, have been found which show increased levels of activity for stimulation in one wavelength region and a suppression of the resting level of activity in another wavelength region Some

cells may be excited by relatively long wavelengths of light and inhibited by middle wavelengths These are characterized as "red-on, green-off " Other cells may be characterized as 'green-on, red-off " Still other cells have been found which may be characterized as "yellow-on, blue-off" and "blue-on, yellow-off" (Fig 8 90) The precise nature of the response of these cells depends upon the condition of adaptation of the eye

Some effort at determining the spatial organization of the receptive fields of color differentiating cells has already been made by Wiesel and Hubel (1966) They have reported three main types of cells (Fig 8 91), the first of which must be involved in both spatial and color discriminations, the second of which is apparently involved primarily in color discrimination and the third of which is apparently specialized primarily for a role in spatial discrimination

The first class of cells has either an excitatory or inhibitory center with an opposing surround Center and surround have different spectral response properties Receptive field sizes are relatively small, many well under 1° in diameter The on-center types are smaller than the off-center types and all tend to be smaller when their receptive fields are located closer to the fovea In order of frequency, they have been found with "red-on" centers and "green-off" surrounds, "red-off" centers and "green-on" surrounds, "green-on" centers and "red-off" surrounds, "green-off" centers and "red-on" surrounds, and "blue-on" centers and "green-off" surrounds Other possible color combinations were not found

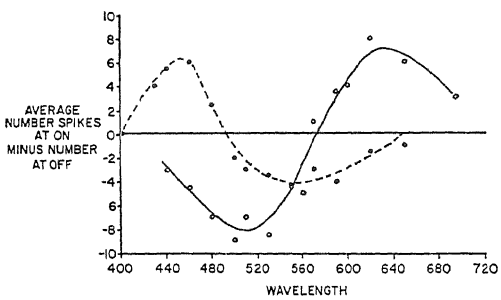


Fig 8 90 Difference between average number of spike discharges at stimulus onset and average number of spike discharges at stimulus offset as a function of wavelength (nm) of the stimulus for each of two single neural units in the lateral geniculate body of the monkey The results illustrated with solid circles and a dashed line show a maximum *on* response in the blue region and a maximum *off* response in the yellow region Results presented with open circles and a solid line indicate a maximum *off* response in the green region and a maximum *on* response in the red region of the spectrum (DeValois 1960)

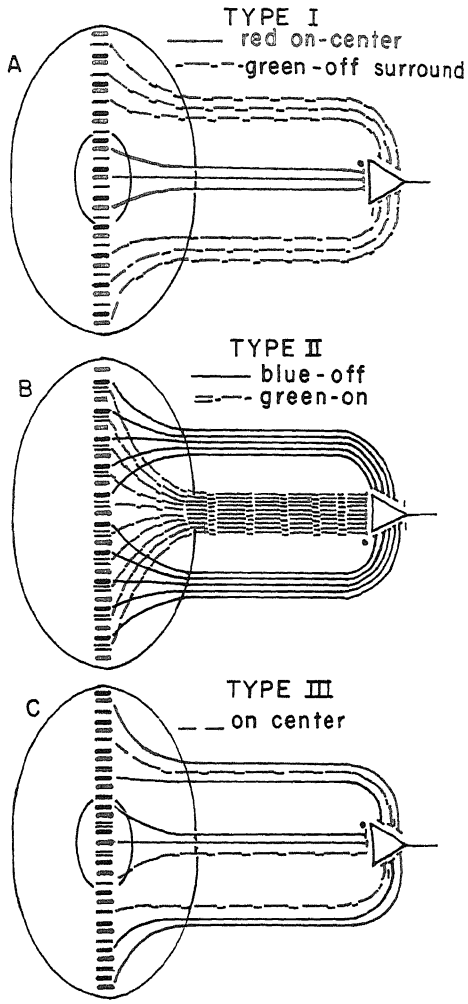


Fig 8 91 Nature of units showing a selective response to wavelength of stimulus in the lateral geniculate body (Wiesel and Hubel 1966)

The second type of cell, a small minority of the total, showed no differential center-surround organization Opposing responses dependent upon the wavelength of stimulation were found over the entire receptive field, sometimes "green-on, blue-off" and sometimes "green-off, blue-on " A very few were found with red and green connections These receptive fields also tended to be less than 1° in diameter

The third type of cell was found to have a concentric "on" or "off" center and opposing surround but no spectral differentiation between center and surround Still another type of cell was found in the ventral layer of the LGN with a concentric "on" center and a very large "off" surround The spectral response properties of these cells tend to be displaced toward the long wavelength end of the spectrum When both center and surround were stimulated, the response characteristic of the surround tended to prevail The opposite is true for concentrically organized receptive fields of the retina and most of those reported heretofore in the LGN

THE ACTIVITY OF SINGLE CELLS OF THE CORTEX

Receptive Fields. Unit activity of cortical cells has been studied by a large number of investigators, but the work of Hubel and Wiesel (1962 to 1965) has probably attracted the most attention In most of their work, these investigators have found almost exclusively elongated receptive fields associated with single cells when a spot of light was employed for mapping of the fields Receptive field organizations are shown in Figure 8 92 for comparison with two receptive fields found in the lateral geniculate body The linear organization of the cortical receptive fields found in Area 17 as illustrated in the figure has been characterized as "simple " In Area 17 and more often in Areas 18 and 19, receptive field organizations are found of a more complex sort These also respond optimally to stimulation by a line element, either a white line on a dark field or the converse

For both simple or complex cells, it is important that the orientation of the line correspond to the optimum for the receptive field The complex cells respond to stimulation which is relatively independent of the precise position of the stimulus, however Orientation remains critical but the line element may be displaced laterally Higher order hyper-complex cells have been recorded for which the length of a line element stimulus may be important as well as other geometric characteristics of the stimulus

Presumably, ganglion cells, the receptive fields of which represent linear arrays on the retina, each project to individual cells of the

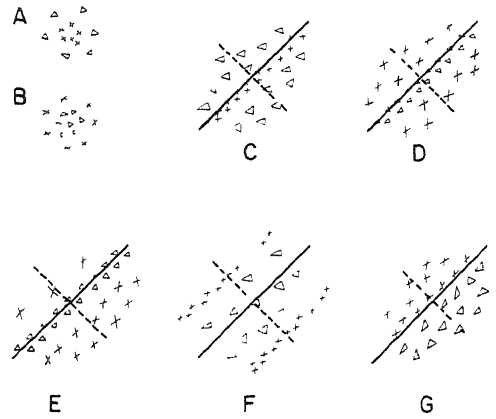


Fig. 8 92 Receptive fields of simple cells in the cortex of the cat compared with the circular organization of receptive fields of the lateral geniculate body *A on-center geniculate receptive field B off center geniculate receptive field C to G* various arrangements of simple cortical receptive fields Areas indicated by × produce an *on* response with retinal stimulation areas indicated by △ are inhibited by retinal stimulation and discharge when the stimulus is terminated A solid line shows the axis of orientation of the receptive fields Each kind of arrangement is found in all orientations (Hubel and Wiesel 1962)

lateral geniculate body, all of which in turn converge upon a single cell in the striate cortex Such single cells then show the kind of organization illustrated in Figure 8 92 Sets of these simple cells which represent parallel line elements must then converge upon single cells which show the complex field organization Higher orders of complexity are associated with higher orders of convergence of cortical cells This kind of arrangement is illustrated in Figure 8 93 Some of the work of Hubel and Wiesel has been confirmed by other investigators (Baumgartner, Brown, and Schulz, 1965, Spinelli and Barrett, 1969) but these other investigators have also found some cells in the cortex with circular receptive field organizations

As they reported earlier for the cat (1962), Hubel and Wiesel (1968) have reported columnar organization of cells in the monkey cortex Columns of cells oriented perpendicularly to the cortical surface contain cells which all have similar receptive field orientations An independent and overlapping set of somewhat larger columns contains cells which tend to respond to one or the other eye There is some indication that cells may also be grouped on the basis of their responses to movement

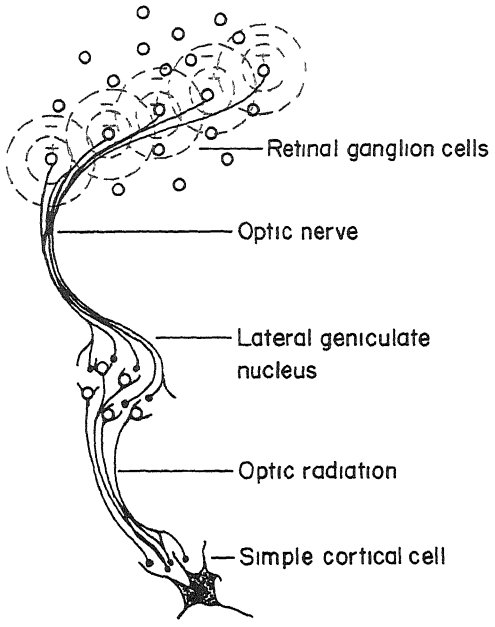


Fig 8 93 Connections from retinal ganglion cells through cells of the lateral geniculate body to the single cells of the cortex which respond to line element stimulation of the retina

Some cells in both cat and monkey show responses which are specific to the direction of motion of the visual stimulus In monkeys, receptive fields tend to be smaller and sensitivity to stimulus pattern orientation more specific than in the cat

In addition to the columnar organizations, there appears to be organization in depth in the visual cortex In the monkey, complex and hypercomplex cells which are driven by either eye are found in the outer layers (II and part of III) In middle layers (III and IV) simple cells predominate In layer IV, units lacking in orientation specificity have been found in the cat, as reported by Baumgartner and co-workers (1965) and Spinelli and Barrett (1969), but these units could conceivably be the axons of geniculate cells In the monkey, most of the cells of layer IV respond only to stimulation of one eye The cells of layers V and VI, like those in layers II and the upper part of III, have mostly complex and hypercomplex fields and are binocularly driven There are relatively few binocularly activated cells in Area 17 of the monkey Binocularly activated cells are common in this area in the cat, on the other hand.

Color Coding. Only a small percentage of the cells of the cortex of the monkey which have been examined for their possible color coding characteristics have proven to be responsive to differences in the spectral character of the stimulating light (Hubel and Wiesel, 1968) Some of these were found to have concentric field organization unlike any found in the lateral geniculate body Only 25% of the simple cells and 7% of the complex and hypercomplex cells of the cortex gave responses specific to color of the stimulus The nature of color coding in the cortex has been studied by other investigators, but there is no clear agreement as to its nature (Motokowa, Taira, and Okuda, 1962, Gouras, 1970)

Binocular Coding. As indicated earlier, the first significant interaction of neural activity from the two eyes probably occurs in the cortex Recently, receptive fields have been found in the cortex which are driven more effectively by stimulation of both eyes than by stimulation of either eye alone The locations of the receptive fields on the retinas of the two eyes are not always in the same relation to the area centralis of the cat, however, or in the same relation to the fovea in the monkey The field of one eye may be slightly further away from the center than that of the other Thus, when the retinal center is being stimulated by a given point in space, some other point or object in space must be either further away or closer to the animal than the point fixated in order to stimulate this particular kind of receptive field equally well in both eyes This is due to the disparity of image location in the two eyes for any objects in the visual field which are at a different depth than objects which stimulate corresponding points (Barlow, Blakemore, and Pettigrew, 1967) When the eyes of the cat are immobilized, an optimum distance can be found for a line element stimulus in the proper orientation which will give rise to maximum response Presumably, this distance is such that for the orientation of the two eyes the stimulus object will be imaged on both of the receptive field locations for the particular cell from which recordings are being made An illustration of the typical disparity found, on the average, is presented in Figure 8 94 for both the vertical and the horizontal direction It is evident that the range of disparity is much greater for the horizontal than for the vertical direction When the head of the animal is in a normal erect position, the two

eyes are displaced horizontally, and therefore, disparities resulting from differences in depth of objects in the visual world are horizontal In the monkey, unlike the cat, cells which respond optimally when both eyes are stimulated have only been found in Area 18 About half of the cells in Area 18 require stimulation of both eyes for the elicitation of more than a weak response For some of these, the receptive fields of the two eyes correspond geometrically, for others there is a disparity in positions of the two fields (Hubel and Wiesel, 1970)

Work of this kind suggests that specific cortical cells which respond to objects in the visual world at specific distances relative to the point of regard of the organism may be involved in the process of binocular depth discrimination They could also be involved in the control of vergence movements of the eyes associated with changes in depth of fixation

DEVELOPMENTAL ASPECTS OF CORTICAL PROCESSES

In recent years, there has been considerable interest in the relation of visual experience to the development of the corticoneural processes which subserve vision Experiments have been performed on cats in which several different

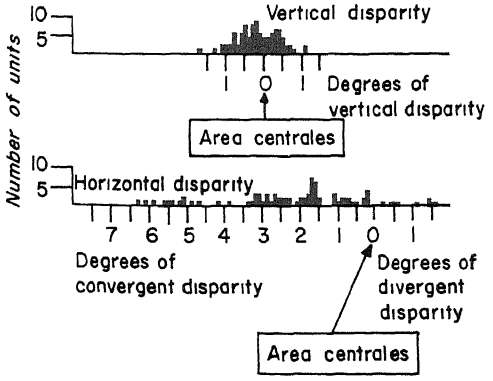


Fig 8 94 Histograms of horizontal and vertical disparities of the centers of 87 binocularly stimulated units in seven cats Estimated positions of the two areas centrales were assigned zero disparity Horizontal angular separation of the two receptive field centers less than that of the areas centrales were taken to indicate convergent disparity greater angular separations were taken to indicate divergent disparity Units with convergent disparity are excited optimally by objects located closer to the animal than its point of fixation The reverse is true for units with divergent disparity (Barlow, Blakemore, and Pettigrew, 1967)

procedures have been employed (Ganz, 1971) If both eyes are covered at birth or sutured shut and left that way for the first 3 months of life, there is a gross impairment of vision Subsequent electrophysiological studies indicate that fewer cortical cells with elongated receptive field organizations are found Whereas in the normal animal, virtually all cortical cells are responsive to visual stimulation, more than a third of the cells in the binocularly deprived animal are unresponsive (Wiesel and Hubel, 1965) Those cells which do respond show a proportion responsive to stimulation of either eye which is approximately the same as that in the normal animal

If only one eye is kept closed for 3 months, the effect is very much more severe for the deprived eye than is the case with binocular deprivation Wiesel and Hubel (1965) found that only 7% of cortical cells from which they recorded would respond to stimulation of the deprived eye

Neurons in the visual cortex of kittens 8 to 16 days of age respond to complex visual stimuli such as edges and contours in a way which is qualitatively similar to the response found in normal adult animals (Hubel and Wiesel, 1963) This suggests that the effect of deprivation is some kind of degenerative process rather than a failure of development in the absence of stimulation Clearly, the degeneration, whatever its nature, is more severe when an alternative mechanism, i e , an uncovered eye, is available

Behavioral studies with cats which have been visually deprived confirm the neurophysiological studies Binocularly deprived animals have difficulty in making form discriminations which are relatively easy for the normal animal For an animal which has been monocularly deprived for the first three months, such discriminations may be impossible with the deprived eye Brightness discrimination, color discrimination in primates, and gross form discriminations may show little or no effect of binocular visual deprivation (Riesen, 1958) Impairment of the deprived eye is noted after protracted monocular deprivation, however, for brightness discrimination and gross form discrimination The impairment would appear to be fairly specific to finer kinds of pattern vision after binocular deprivation

In the normal animal a high percentage of the cortical cells examined may be stimulated

via either eye Both eyes may participate in excitation of a large majority of the cells If, at birth, the ocular muscles are severed so that conjugate eye movements are impossible, a different result is found The elongated receptive fields of the cortex can be detected, but the majority of them are stimulated either by one eye or the other, not by both eyes (Hubel and Wiesel, 1965)

The shorter the duration of the interval for which eyes are closed, the less marked are impairment effects on vision There does appear to be a critical duration which, in the case of the cat, is approximately the first three months after birth Extensions of this work with the monkey indicate that the most critical period for deprivation of vision is in the first 6 weeks of life, but that the critical period lasts for at least the first year of life

CONTROL OF EYE MOVEMENTS

Movements of the eyes are extremely complex Four distinct types of movement may be distinguished As we look about us, without moving our heads, at stationary objects within the visual field, our eyes move in their sockets in a conjugate fashion so that both are always directed toward the same point of regard Such eye movements are called saccades They are extremely rapid and may reach a peak velocity of 400°/sec (Robinson, 1968) The reaction time for a saccadic movement is approximately 200 msec These movements are classified as voluntary, but they may be initiated involuntarily by peripheral visual stimulation They may also result from a lateral auditory stimulus Neural control of these movements appears to be coded in terms of eye position rather than eye movement (Robinson, 1970, Schiller, 1970)

A second class of eye movement is the smooth pursuit movement This is only possible when the eye is tracking a moving target All other voluntary movements are saccadic When some object moves within the visual field, it is possible for the eye to pursue it smoothly Such movements may be voluntary or involuntary If a repetitive pattern is moved continuously in front of the eyes, particularly in a horizontal direction, it is virtually impossible to prevent pursuit movements by the eyes Such movements are referred to as the *optokinetic reflex* Control of pursuit movements as contrasted with saccadic movements would appear to be based on velocity rather than position of the stimulus The eyes are capable of following a moving target with reasonable accuracy up to approximately 30°/sec Smooth pursuit movements up to velocities as high as 100°/sec are possible but not in maintained synchrony with a moving target (Robinson, 1968)

Both saccadic movements and pursuit movements are conjugate in that the two eyes move together as the point of regard is shifted in a frontal plane On the other hand, when fixation shifts from one object to an object which is located at a different depth, vergence movements occur If we look from a near to a far object, the eyes must diverge in order to maintain fixation With transition of the point of fixation to an object much closer, the movements will include convergence

The visual world appears to be remarkably stable in view of the mobility of our heads Compensatory eye movements which provide correction for movements of the head, as long as these are not excessive, are controlled by the semicircular canals of the vestibular system These represent a fourth class of eye movements When movements of the head are extreme, such as is the case when the body is rotated continuously about a vertical axis passing through the head, eye movements resulting from stimulation of the semicircular canals may be maladaptive Oculovestibular nystagmus occurs with a relatively slow motion of the eyes in the direction opposite that of rotation and a fast motion in the direction of rotation The slow phase is apparently controlled by the vestibular mechanism and is compensatory in nature When the eyes reach their anatomical limit, the fast phase of nystagmus serves to "reset" them for a subsequent repetition of the slow phase Slow and fast phases of motion alternate as long as the eye is exposed to continuing angular acceleration The slow phase is initiated by the vestibular mechanism while the fast phase is apparently under control of the same system which is involved in saccadic movement While smooth pursuit movements can follow stimuli at a rate of up to only 30°/sec , the vestibular system can compensate for angular rotations of greater than 300°/sec This would seem more than adequate for most rotations of the head, but it is not sufficient for such extreme rates of rotation as those which may be undergone by a

figure skater executing a spin Rates of over a 1000°/sec have been measured in these circumstances

Independence of the neural mechanisms involved in smooth pursuit movements and saccadic movements is illustrated by the differences in appropriate visual stimuli for their elicitation (Rashbass, 1961) The separation of the systems is further supported by the fact that they are influenced differentially by the use of barbiturates (Rashbass and Russell, 1961) These drugs act on the smooth pursuit system before any influence can be detected on the saccadic system Similarly, vergence eye movements have been demonstrated to be distinct from conjugate movements (Rashbass and Westheimer, 1961) They, too, show differential sensitivity to barbiturates (Westheimer and Rashbass, 1961)

Each of the four systems for eye movement control converges upon the motor nuclei that innervate the extraocular muscles with resulting movements of the eyeball (Fig 8 95) Movements induced via the various systems are additive at this final common path

CENTRAL NERVOUS SYSTEM INVOLVEMENT IN EYE MOVEMENTS

The Saccadic System. Saccadic movements of the eyes may be evoked by stimulation of Area 8 of the cortex, the frontal eye fields (Fig 8 96) Conjugate contralateral saccades are induced by a brief stimulus train in the unanaesthetized monkey (Robinson and Fuchs, 1969) These movements cannot be distinguished from the animal's spontaneous saccades Extent and direction of the motion appears to depend upon the region of the frontal eye fields stimulated Once threshold for the effect has been exceeded, the movements occur in an all or none fashion ranging from 2° to 60° in extent with amplitude as well as direction dependent upon stimulus location Experimental investigations of the result of artificial stimulation of the frontal eye fields would appear to indicate that they are adequate for the control of all saccadic movement of the eyes Fibers from the area descend in the cortical bulbar tract to the pontine reticular formation

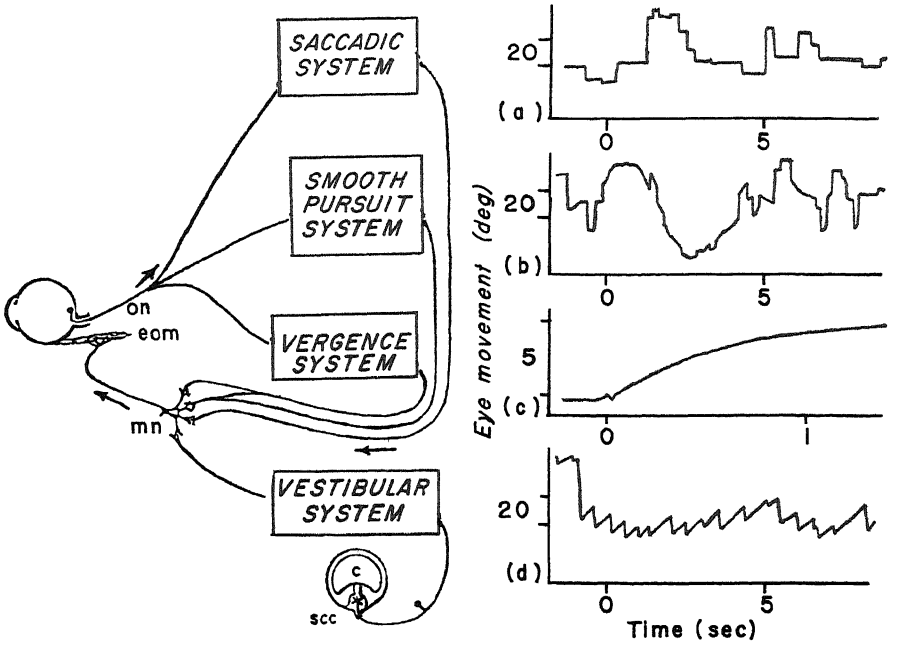


Fig 8 95 The primate eye movement control system The types of eye movements controlled by the saccadic system are illustrated at a the smooth pursuit system at b the vergence system at c and the vestibular system at d Each system converges on the motor nuclei (mn) which innervate the extraocular muscles (eom) and move the eyeball Optic nerve on cupula c semicircular canal scc (Robinson 1968)

Unfortunately, the attractive conclusion that saccadic movements are initiated by frontal eye fields is rendered difficult by the fact that recordings from single units in this area indicate that such units fire *during* the occurence of saccades rather than in advance of their occurrence (Bizzi, 1968, Bizzi and Schiller, 1970)

The Smooth Pursuit System. Smooth pursuit movements are believed to depend upon information from Areas 17, 18, and 19 of the visual cortex (Fig 8 96), as well as from the superior colliculi Cells in these regions may provide information as to the direction and velocity of image movement on the retina (Robinson, 1968) The tegmental region is probably also involved Velocity information from the cortex presumably joins information from the superior colliculi and the tegmental and pretectal regions in the corticotectal tract The role of the superior colliculi is somewhat in doubt, and it has been demonstrated that eye movements may occur in the absence of these structures (Pasik and Pasik, 1964) Goal directed movements of the eyes, i e , changes of fixation to specific points in the visual field, may involve the superior colliculus Stimulation of specific points in the superior colliculus has been shown to result in eye movements of a highly specific sort (Apter, 1946)

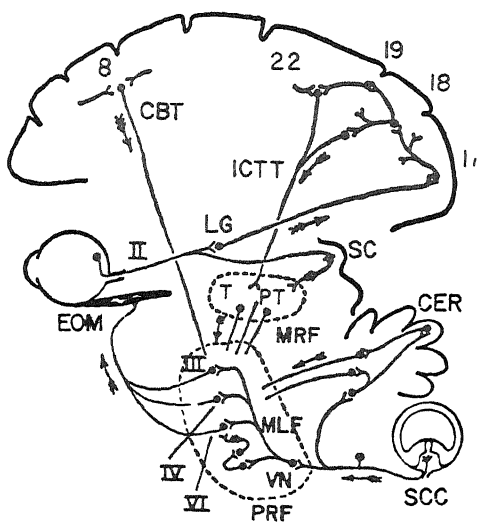


Fig 8 96 Major oculomotor pathways in the central nervous system CBT corticobulbar tract CER cerebellum EOM extraocular muscles ICTT internal corticotectal tract LG lateral geniculate body MLF medial longitudinal fasciculus MRF mesencephalic reticular formation, PRF pontine reticular formation PT, pretectal nuclei SC superior colliculi SCC semicircular canals T, tegmentum VN vestibular nuclei, II optic nerve, III IV VI, oculomotor trochlear and abducens nuclei and cranial nerves, 17 18 19 22 Brodmann s areas of occipital and parietal visual and association areas, 8 Brodmann's premotor area in the frontal lobes (Robinson 1968)

Vergence Movements. When cortical Areas 19 and 22 are stimulated electrically in the lightly anesthetized monkey, vergence movements result Presumably, information in these cortical areas is transmitted via the internal corticotectal tract to the brain stem (Robinson, 1968)

The Vestibular System. Pathways involved in the mediation of ocular response to vestibular stimulation are perhaps the best known of all those involved in eye movement control Signals from vestibular nuceli are relayed along the medial longitudinal fasiculus to the appropriate motor nuclei and thence to the extraocular muscles Diverging pathways extend to the pontine reticular formation and the cerebellum as well As mentioned above, the fast phase of vestibular nystagmus may be under control of those pathways which mediate saccadic movements In any case, the fast phase depends upon the brainstem between the vestibular and oculomotor nuclei

Robinson (1968) has pointed out that the mesencephalic and pontine reticular formation are extremely important in eye movement control All four of the oculomotor subsystems illustrated in Figure 8 96 convey information to these formations, and it may be here that the final integration of activity in each of these systems is accomplished prior to the transmission of signals to the final common path

BASIC VISUAL FUNCTIONS

The rich visual experience which our eyes afford us may be understood in terms of several fundamental response properties of the visual system One of the most primitive is the ability of the eye to discriminate differences in brightness This is a fundamental ability and may have been the primary function of evolutionary precursors of the complex vertebrate eye The vertebrate eye has the capability of differential brightness discrimination over the area of the retina This spatial discrimination capability is perhaps the most interesting and most important characteristic of the visual system Its utilization depends in large measure on the ability of the visual system to make temporal discriminations as well Very

rapid changes in the character of stimulation of a given point on the retina or of the whole pattern of stimulation of the retina may be followed without difficulty The interval in which the eye is capable of integrating energy from external sources of light represents a useful compromise The integrating capacity permits the utilization of lower flux levels for seeing On the other hand, if integration occurred over too long a period, the effect would be to blur the visual world as effects of one stimulus merged with those of another The integration interval is matched nicely with the rates of eye movement and the average duration of fixation on a given point in the visual world

The range of illuminations over which the eye is capable of functioning is extremely great From the lowest threshold luminance to which the eye can respond to the maximum luminance which can be viewed without damage to the eye, there is a range of over 1,000,000,000 to 1 Nevertheless, the eye does not function efficiently immediately when the illumination level is changed Although it is capable of adapting fairly rapidly from low levels to high levels, it requires a fairly long time interval to adapt from relatively high levels of illumination down to the lowest levels which can be detected In the latter case, a half an hour or longer may be required

Another fundamental capability of the normal human eye is the differential discrimination of light based on its wavelength composition This is a function which involves the cones and relatively high levels of illumination

Finally, the eye has the ability to make spatial resolutions in depth as well as in the projections on the plane of the retina Of the variety of cues for the discrimination of depth, one of the most important is the disparity of the elements which are illuminated by corresponding parts of images on the two retinas by reason of the fact that the two eyes are separated in space as they view the three-dimensional world in which we live

Functions Related to Brightness

BRIGHTNESS DISCRIMINATION

Innumerable experiments have been performed to ascertain the relations among the amount of light required for detection and the luminance of the background, the location of the light in the visual field, the size of its image on the retina, its temporal duration, and its spectral characteristic and the spectral characteristic of any background illumination which may be present as well as the condition of adaptation of the eye (Graham, 1965a) When the purpose is to determine the minimum luminance which can be detected either against a dark background or an illuminated background, such experiments are classified as brightness discrimination experiments (Brown and Mueller, 1965) For the observer, the discrimination is always relative to the spatial or the temporal dimension A region of the visual field is seen as brighter than its surround or a region of the visual field (perhaps the entire visual field) is observed to become brighter at some point in time relative to its previous appearance Such experiments may also be carried out with a decrement of luminance rather than an increment when a sustained adapting level of illumination is employed (e g , Herrick, 1956)

The experiment reported earlier on the absolute minimum amount of light energy which can be detected (Hecht, et al , 1942) represents an extreme on the continuum of brightness discrimination experiments As the eye is adapted to higher and higher levels of illumination, the luminance of an increment which can just be detected against such a background adaptation level increases in accordance with the function represented in Figure 8 97 It is evident that the required increment in retinal illumination increases at an increasing rate as the background luminance is increased Presumably, there is some level of background illumination sufficiently high that no additional increment can be discriminated against it Such a background level would completely saturate the visual system

The data in Figure 8 97 represent the test intensity required for the detection of a light against a background adapting intensity for a human observer and in the retina of the monkey Data from the retina of the monkey are based on a selected criterion response level of the late receptor potential (Brown, 1968), this is recorded with a tiny electrode inserted into the retina at an appropriate depth to provide an index of the activities of the receptors themselves The similarity between the two functions is rather striking It suggests that the form of the intensity discrimination response

function may be determined in the receptors themselves It should be recognized, however, that the late receptor potential as recorded represents the activity of many receptors It is undoubtedly also influenced by interactions which occur as a result of connections of the horizontal cells The function which represents the data for the monkey retina in Figure 8 97 must therefore be recognized as representative of the retina but perhaps not exclusively of the receptors themselves When the eye is adapted to a higher luminance, a significantly higher test stimulus is required for detection For any given level of adaptation the nature of the response at the retina is remarkably similar to that at other adapting luminances In Figure 8 98, the response in terms of percentage of the maximum late receptor potential is plotted as a function of the test intensity for each of six conditions of adaptation The functions, with the exception of that for dark adaptation, are remarkably similar and are simply shifted to higher ranges of test intensity with increase in the level of the adapting intensity (Boynton and Whitten, 1970) The range of response of the system thus remains approximately the same for any condition of adaptation, but as the adapting conditions are elevated, there is an expansion of the range of luminances over which gradations in response may be obtained The "gain" of the system is thus reduced as higher levels of background luminance are encountered and this occurs in the early processes which take place within the retina itself This kind of process is undoubtedly involved in the ability of the eye to function over an intensity range of greater than 10^9 1

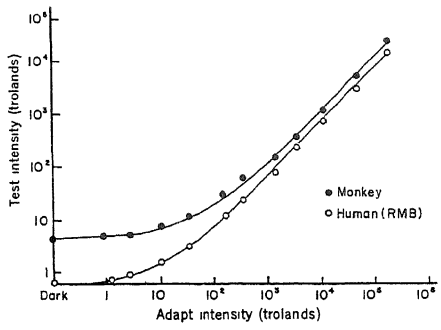


Fig 8 97 Log of threshold retinal illumination (trolands) as a function of the adapting level The solid points represent a 10 μv criterion response level of the late receptor potential of a monkey The open circles are based on human foveal thresholds measured with the same apparatus (Boynton and Whitten 1970)

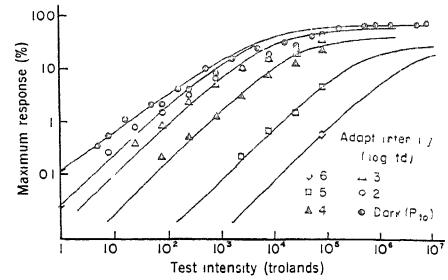


Fig 8 98 Late receptor potential amplitude (percent of maximum) as a function of test flash intensity for each of five adaptation intensities and for dark adaptation Average values for six monkeys (Boynton and Whittel 1970)

DARK ADAPTATION

There are important changes which occur within the visual system when illumination is reduced from daylight levels to very low levels before it is possible for even the crude low luminance or "scotopic" vision to become optimum The process of "dark adaptation" requires a half an hour or longer after an abrupt transition from a high illumination surrounding to one of relative darkness (Bartlett, 1965b) In Figure 8 99, the minimum amount of light which can just be detected is represented on the ordinate and time in the dark is represented on the abscissa The ordinate axis is plotted on a logarithmic scale, thus greatly amplifying the lower portion of this relationship For a period of from 8 to 12 min after a high luminance adapting field has been removed, the threshold value at first reduces rapidly and then the rate of reduction in threshold is slowed, reaching a minimum at about 10 min A little later there is a subsequent additional reduction in threshold which again occurs at a decreasing rate and approaches a minimum at some time after 30 min in the dark

For various reasons, the higher threshold portion of the curve in Figure 8 99 has been associated with the cones and the lower threshold portion has been associated with rods In the first place, when a small test stimulus is employed which is restricted to the fovea, two branches are not obtained In the second place, if a red test light is used for which the rods are no more sensitive than the cones, the second branch of the curve does not appear, even though the test light may be presented to a region of the retina which is rich in rods If a subject in a dark adaptation experi-

ment is asked to identify the color of a test flash, he is able to do so during most of the time for the first 8 or 10 min of dark adaptation, corresponding to the initial segment of the curve in Figure 8 99, but is unable to do so unless the test light is restricted to extremely long wavelengths at later times during dark adaptation Thus, color discrimination is approximately associated with the "cone" portion of the curve and the inability to discriminate color is associated with the "rod" portion of the curve

It was believed for many years that the time course of dark adaptation could be understood in terms of the time required for regeneration of the photopigments of the visual receptors which are bleached by any high level adapting light (Hecht, 1937) The time course of dark adaptation certainly is associated with the regeneration of photopigments, but the level of threshold cannot be explained directly in terms of the concentration of photopigment present or the quantum catching capability of the retinal pigments It has been demonstrated fairly conclusively (Rushton, 1962) that the threshold is still very much elevated even after 90% of the photosensitive material has been regenerated following bleaching of substantial percentages of the photopigment

The time course of dark adaptation, as suggested above, is greatly influenced by the testing procedures employed This is illustrated in Figure 8 100, which shows three different dark adaptation curves obtained with each of two test field sizes in two different retinal regions (Arden and Weale, 1954) When a small test field is confined to the foveal center of the eye, a single function is obtained which shows the reduction in threshold at a decreasing rate with time If the same small test field is presented at a location 8° away from the foveal center, a curve of a different sort is obtained There is an initial function showing reduction in threshold at a decreasing rate followed by a subsequent similar function which carries threshold down to a lower level which,

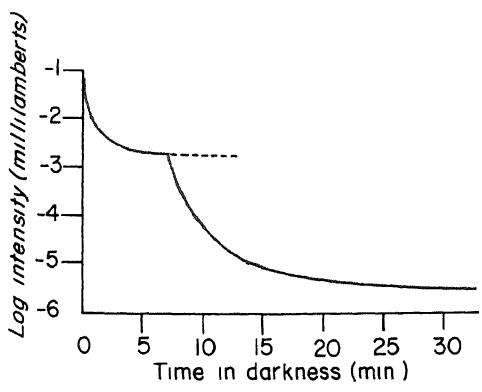


Fig 8 99 The time course of reduction in a light detection threshold during dark adaptation The initial branch of the curve is associated with photopic vision and is attributed to cone function The later branch is often called the scotopic branch and is associated with rod function

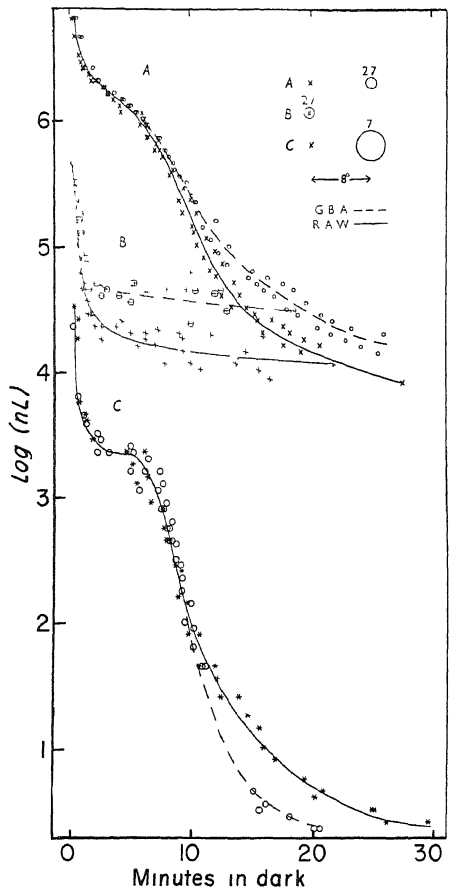


Fig 8 100 Dark adaptation curves measured for two test field areas and two retinal locations Results are presented for each of two subjects for each of the three conditions The upper functions represent a very small test spot centered on the fovea or located 8° from the fovea The lower functions represent a large test area centered 8° from the fovea The larger drop as compared on a logarithmic axis of the late branch of the function for the large test area with the smaller test area in the periphery indicates a larger relative change in threshold with the large test area This suggests a spatial summation effect which is developing during the time interval represented by the function (Arden and Weale 1954)

in fact, corresponds fairly closely with the final level reached in the fovea If a substantially larger test field is employed centered on a point 8° away from the fovea, again a curve of two branches is obtained, but the function is displaced downward The initial branch is very similar to that obtained with a small test field although displaced downward on the threshold axis The later branch, however, is much steeper than that obtained with the small test field Here, as in Figure 8 99, threshold is presented on a logarithmic axis The significance of this is that a given distance along the threshold axis represents relative amount of change quite independent of the location along the axis Thus, there is a relatively much greater amount of change in threshold in the same time interval for the later branch of the dark adaptation curve when a large test field is employed The function can thus hardly be described as exclusively dependent upon the regeneration of photopigment Relative change in threshold would be uninfluenced by such a variable as test field size if the sole determinant of threshold were the relative amount of photopigment available The much steeper "rod" branch for the larger test field in the periphery can most probably be explained in terms of some sort of increased capability of the eye to summate stimulus energy over area during dark adaptation There are other evidences for such a change in addition to the data presented in Figure 8 100 Direct tests of the ability of the eye to summate stimulus energy over area show that this can be accomplished to a greater degree in the dark-adapted as compared with the light-adapted state (Bartlett, 1965b)

The data of Figures 8 99 and 8 100 show that when the eye is placed in darkness it becomes more sensitive by a factor of 1,000,000 1 or greater as it remains in darkness for up to 30 min or longer This has been attributed, as suggested above, to increased ability of the retina to summate energy over area There are, in fact, three mechanisms involved, but the increased areal summation capability is probably the most important First, when the eye is placed in darkness, there is an increase in the size of the pupil The pupil may expand in diameter from a minimum of little more than 2 mm to a maximum of something in excess of 8 mm. This represents a range of change in area by a factor of 16 1 To that extent, the eye has become capable of gathering light more effectively from any given external light source, but the increase in effectiveness is hardly sufficient to explain the tremendous range of change which accompanies dark adaptation Secondly, the bleached photopigments do regenerate, but as Rushton has shown, the logarithm of change in threshold is more nearly proportional to the change in concentration of photosensitive material than is the change in threshold itself The tremendous range of change during adaptation is most readily explained in terms of changes in retinal organization which permit increased summation of energy over area and time as well as to some mechanism which increases the responsiveness or "gain" of the system at very low levels of illumination (see below) Such a mechanism is implied by the parallel functions of Figure 8 98 From these it is clear that much smaller ranges of change in stimulus energy result in the same range of change in response in the retina when the eye is adapted to very low luminances There is a variety of theories as to the precise mechanisms involved (Barlow, 1964, Dowling, 1967, 1970)

Barlow has suggested that the mechanism of dark adaptation may involve the gradual subsiding of a retinal noise level which is the result of photic stimulation According to his interpretation, following termination of a light stimulus, the receptors themselves maintain some level of activity which is reduced only gradually Dowling has shown that the B-wave of the electroretinogram shows a relation to increased background luminance which is similar to that shown by the activity of retinal ganglion cells The ganglion cells in turn are excited by the bipolar cells either directly or via amacrine cells Dowling therefore speculated that the adaptation process may involve some kind of inhibitory feedback of amacrine cells on bipolar cells which serves to reduce the gain of the system at high levels of illumination More recently (1970) he has shown that the B-wave of the electroretinogram as measured in a gross retinal response is most nearly duplicated by the response of glial cells when the response characteristics of individual components in the retinal complex are compared

It should be evident that the precise mechanisms which underlie the adaptability of the visual system to such a broad range of condi-

tions of illumination are as yet unknown It has been clearly demonstrated, however, that with increasing dark adaptation there is increased spatial and temporal summation at the retinal level These changes permit more efficient utilization of small quantities of energy They occur at the expense of both spatial and temporal resolving power of the eye as discussed further, below

There are certain misconceptions concerning the effect on vision of adaptation to very low luminances The increased temporal and spatial summation capabilities permit detection and utilization of low light levels They do not preclude fine spatial and temporal resolution if sufficient light is presented to the eye when it is in a dark-adapted state That is, the dark-adapted eye is capable of just as high a level of spatial or temporal acuity as is the light-adapted eye, given sufficient luminance to work with Of course, if the necessary luminance remains available for any length of time, then the state of adaptation is changed from one of dark adaptation to some level of light adaptation Nonetheless, fine spatial detail can be discriminated quite well in a very brief flash of sufficient luminance presented to the dark-adapted eye (Brown et al , 1953) This is illustrated in Figure 8 101, which shows the time course of dark adaptation measured with the presentation of brief test flashes for a variety of visual tasks If the eye is required merely to detect the presence of a light, then the lowest curve is obtained If the eye is required to identify the orientation of a line pattern in the test flash, then the progressively higher curves are obtained with the finest line pattern resulting in the highest of the curves It is evident that for a line pattern which represents a sufficiently small image size on the retina, there is no late rod branch of the dark adaptation curve Clearly, the rods are unable to make a contribution in a visual task of this sort and the function is performed by the cones The dark adaptation of the eye, however, does not interfere with the identification of fine patterns It might be supposed that the increased sensitivity of the rod receptors which are themselves incapable of resolving fine detail might somehow interfere with the resolu-

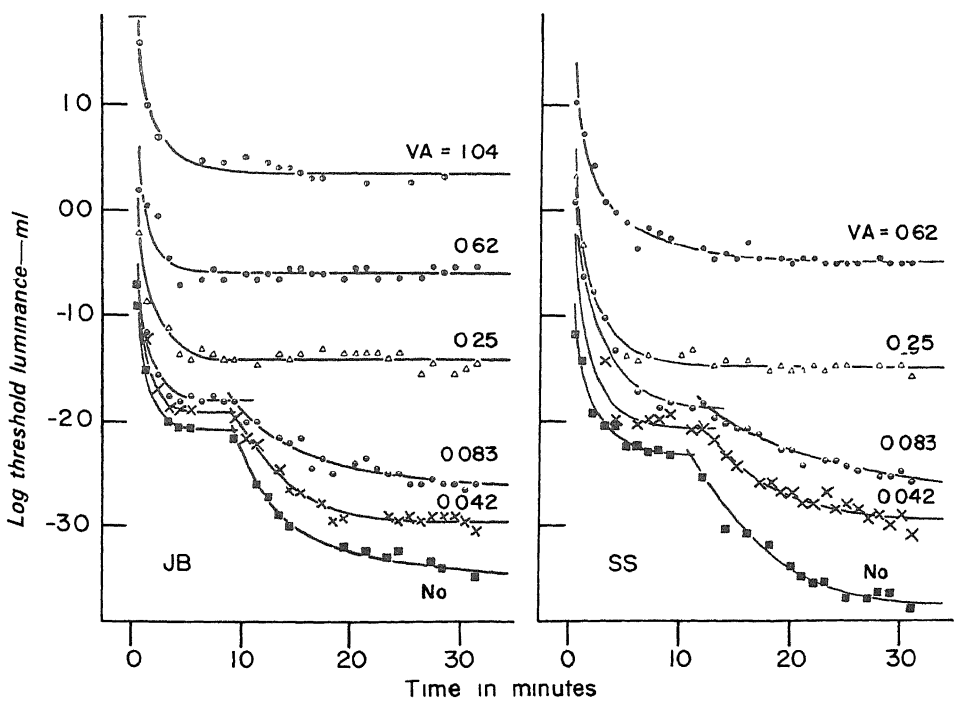


Fig 8 101 Luminance thresholds for the resolution of acuity gratings during dark adaptation for each of two observers The number beside each curve refers to the level of acuity (reciprocal of the visual angle subtended by a grating line) The bottom curve represents simple detection of light with no grating present (Brown et al 1953)

tion of such fine detail by the cones This would be reflected by an elevation in the later portion of the highest curves in Figure 8 101, but no such elevation is observed (Brown et al , 1969)

LIGHT ADAPTATION

Just as the eye adjusts when the visual environment is darkened, so does it also adjust when the environment is made brighter The nature of the response is similar, as illustrated in Figure 8 98, when it is considered on a log log plot, but, as mentioned earlier, this implies a reduction in the gain of the system A given amount of change in response requires a proportionately greater change in the level of stimulation when the adapting level is raised We have seen that it may require more than 30 min for the eye to achieve maximum sensitivity in darkness after it has been exposed to a bright adapting light The time course of adaptation to a lighter environment is usually much more rapid Under most circumstances, light adaptation is complete in from 3 to 5 min , and the eye adjusts to a relatively steady state condition of responsiveness within that time

The time course of light adaptation has been measured in a variety of ways Early measurements were performed by exposing the eye to an adapting light after it had been completely dark adapted and then determining its sensitivity by turning off the adapting light and quickly presenting a test light (Wald and Clark, 1937) Thus, a given adapting luminance would be selected, the dark-adapted eye would then be exposed to this luminance for a short duration, the luminance would be extinguished and at some fixed time very soon after it was extinguished a test flash would be presented for the measurement of threshold This process would have to be repeated time and again with variation in the level of test flash in order to determine a threshold, and again and again for greater and greater durations of light adaptation The results of such studies indicated that with increased duration of exposure to an adapting light, there is an increase in the threshold for detection of subsequent test light as adapting exposures increase up to 3 to 5 min and sometimes much longer The process has been attributed to a decrease in the available photosensitive materials of the retina which continues in time until the rate of regeneration of photosensitive materials sustains a relatively constant and, of course, somewhat lower concentration of photopigments corresponding to the higher level of light adaptation employed

There is more to light adaptation than a change in the average concentration of photosensitive materials, however Light adaptation may be measured in a somewhat different fashion which gives very much different results In the alternative method, a test flash is presented against the light adapting field which is left on The test is thus one of the ability of the eye to discriminate an increment in brightness against the adapting luminance The method is somewhat easier in that test flashes may be presented at various intervals during the entire time course of the light adaptation process On subsequent exposures to the adapting light, slightly higher luminances may be used at a given time when the observer was unable to detect the test flash and lower luminances may be employed when the test flash was visible By repeating the process, it is possible to obtain a measure of threshold for the increment, or test light as a function of duration of the adapting light (Crawford, 1947, Baker, 1949) The results of an experiment performed in this way are presented in Figure 8 102 The function is not a simple one and requires some explanation In the first place, it is important to point out that the experimenter has begun by presenting the test flash in advance of the time at which the light adapting field is turned on This is done in order to provide a baseline level of response in the dark-adapted state A very interesting result of this procedure is the finding that the threshold for detection of the test light begins to be elevated when the test light is presented about 50 msec prior to the presentation of the adapting light Test flash threshold increases rapidly to a maximum at about the time the adapting light is turned on and then reduces in value for a brief interval prior to rising again and reaching a final level which is typical of the specific adapting luminance employed after about 10 min of light adaptation This result may at first seem mysterious in that it appears that the visual system is somehow able to anticipate the occurrence of a stimulus which has not yet been presented This apparent anachronism can readily be understood when we consider that the temporal dimension

in Figure 8 102 is determined by timing devices in the laboratory On the other hand, the actual detection of the test flash is determined within the nervous system Clearly, within the nervous system, influences on a test flash are occurring as the result of the presentation of an adapting light which illuminates the retina as much as 50 msec after the presentation of the test flash The significant factor is that the detection of a flash does not occur at the retina The detection on the part of a conscious observer, as evidenced by his reporting that he sees the flash, probably occurs somewhere in the cortex The relatively low luminance of the test flash initiates a series of events which occupy a finite time interval prior to any detection and that time interval is longer than the time interval required for effects of the adapting flash to influence the same point in the nervous system The latter influence serves to elevate the measured threshold for the test flash

The initial presentation of an adapting light obviously has a significant, albeit transient, effect on sensitivity to a test light This effect is attributed to neural events in the retina Abrupt elevation of retinal illumination probably results in a level of neural activity considerably in excess of that which corresponds to the state of affairs after the adapting light has been on for some time Thus, the system cannot handle increment information as efficiently immediately after exposure to an adapting light as it can after the light has been on for a time The slower rise of threshold for the test light which follows the initial rapid elevation and subsequent decline is attributed to the more gradual change in photochemical balance within the retina

If dark adaptation experiments are performed with a test flash which is presented against the adapting field prior to the beginning of dark adaptation a result similar to that reported above for the initial stage of light adaptation is found to occur (Baker, 1953) There is an elevation in threshold for the test flash which must be attributed to the abrupt change in the level of illumination of the retina with a consequent neural effect which temporarily reduces the ability of the system to handle other additional information This phenomenon is illustrated in Figure 8 103

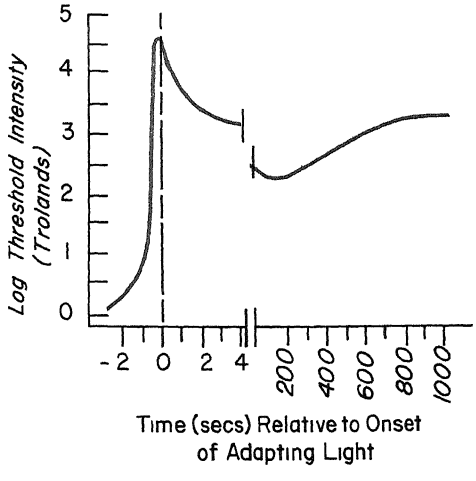


Fig 8 102 The time course of light adaptation as measured by an increment threshold technique Test flash thresholds are measured beginning at a time in advance of the actual presentation of the adapting light and continuing out to approximately 15 min (A hypothetical reconstruction based on the data of Crawford 1947 and Baker 1949)

Discriminations Related to Space or Time

SPATIAL RESOLUTION OF THE VISUAL SYSTEM

As mentioned earlier, one of the most important functions of vision is to provide information as to the relative location and identity of objects in the visual world The retina is extended in space and receives information which differs over its surface in correspondence to the way in which light from objects in the visual world differs spatially There are millions of individual receptor cells distributed over the retinal surface Images of objects in the visual world are distributed over arrays of these cells and the cells then somehow transmit information which the central nervous system is capable of employing for identification of the objects To the conscious observer, there is no complex processing of information required at the conscious level, he immediately apprehends the fact that an object is present and its identy. What he "sees" is for him the "real object" and not some coded neurological process which he must strive to decode

One of the puzzling things about this process is the fact that the visual image quality for the normal observer is better than one might pre-

dict If the aberrations of the optical elements of the eye and the size of the receptor cells are considered, the clarity of vision seems better than it should be Actually, interaction processes within the retina which have been considered above tend to sharpen the process of spatial discrimination and there is not a simple spatial discrimination system based on a point to point correspondence between differences in the level of stimulation of adjacent receptors in the retina and differences in level of excitation of corresponding neurons somewhere higher in the nervous system

Visual Acuity In the case of vision, the term *acuity* is reserved to refer to spatial resolution capability Visual acuity is formally defined as the reciprocal of the minimum visual angle in minutes subtended at the eye by an object in space which can be resolved by the visual system Visual acuity is not an invariant quantity for a given eye or even under a given level of illumination, it depends importantly upon the way in which it is measured

The simplest form of visual acuity is the so-called *minimum detectable* visual acuity. This

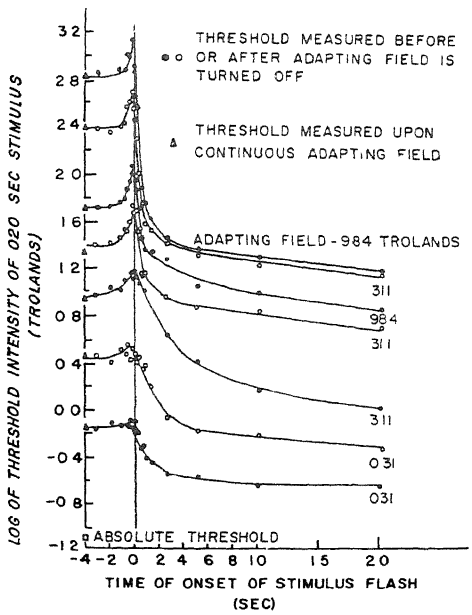


Fig 8 103 Early dark adaptation at 5° from the fovea The adapting field is turned off at time 0 Test flash threshold measurements begin 400 msec prior to the termination of the adapting light and are continued out to 2 sec after termination of the adapting light (From Baker, 1963)

is the discrimination of a small dark spot against a lighter background, or a dark line against a contrasting background Under ideal conditions, the finest line which can be resolved subtends a visual angle of less than 1 sec of arc This represents a visual acuity by the definition given above of greater than 60 The diameter of a dark spot against a lighter background must be somewhat greater then the width of a dark line in order to be discriminated

It is meaningless to attempt to associate a visual acuity value with a light spot or a light line against a dark background Detectability of such a spot or line depends entirely upon the relative luminous flux coming from the line or spot as compared to the background Easily detectable stars are so far away that the visual angles which they subtend at the eye are effectively zero There is sufficient light emanating from them, however, that they can readily be perceived If visual acuity were associated with them by the above definition, it would be infinite in value, hence the concept is meaningless under these circumstances

Another form of visual acuity which is somewhat more complex is sometimes referred to as the *minimum separable* visual acuity This refers to the ability of the eye to discriminate a separation between two points of stimulation This might be referred to two dark spots against a lighter background, or to two light spots against a dark background In this case, it is perfectly valid to consider light against dark and the identification of a separation between two light spots is clearly dependent upon spatial discrimination

Visual acuity measurements made clinically may depend upon a test of the minimum separable threshold The Landolt ring is a commonly used target on charts for the measurement of acuity (Fig 8 104*B*) The gap in the ring is positioned at random from one ring to the next on a given chart and the patient must identify the location of the gap The smallest ring size for which the gap can be correctly located determines the visual acuity with this test For the normal observer, the width of the gap subtends a visual angle of approximately 1 min of arc at the standard measuring distance of 20 ft Normal visual acuity for such a target is therefore *one* It is more often expressed in relation to the distance at which it can be correctly detected for a given observer in relation

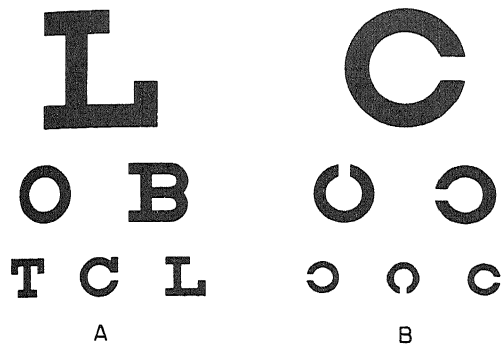


Fig 8 104 *A* Representative letters as used in a Snellen chart for the testing of visual acuity The lines and serifs of the letters are one fifth the height or width of the entire letter *B* Landolt ring test symbols The thickness of the line and the width of the gap are both one-fifth of the outer diameter of the ring (Riggs 1965b)

to a normal population Thus, an acuity of 20/200 implies that a given individual can resolve a test figure at a distance of 20 ft which the normal observer could resolve at 200 ft This represents a severe defect A visual acuity of 20/10 indicates that an observer can resolve a test pattern at 20 ft which the normal observer must view at a distance of 10 ft in order to resolve it

When visual acuity is tested with Snellen letter charts (Fig 8 104*A*), the procedure is very much the same The sizes of those elements which permit discrimination of the letters determine their visibility and visual acuity is calculated from the sizes of these elements for those letters which can just be resolved at the normal distance of 20 ft Again, a visual acuity of one, corresponding to a visual angle subtended by elements of detail of approximately 1 min of arc, is the norm Acuity is expressed as the distance at which the patient can identify the letters in a given line relative to the distance at which a "normal" observer can discriminate the same letters

Visual acuity is sometimes measured with a grating test pattern Such a pattern consists of parallel lines and spaces with lines equal in width to the spaces between them The grating pattern may illuminate a circular patch and may be changed in orientation so that the lines are vertical, horizontal, or oblique The observer is required to identify their orientation One advantage of such a target pattern over individual letters is the fact that the relative illumination in the retinal image is always precisely the same Only the orientation of the lines is altered Here again, the normal value is very close to 1 min of arc for the angle subtended by the width of a single line or a single space at the eye

The level of visual acuity of which the eye is capable is very much limited by the level of illumination available The higher the level of illumination, the higher the visual acuity up to a limiting value for a particular type of target (Shlaer, 1937) In Figure 8 105, visual acuity is seen to increase monotonically with increase in the logarithm of retinal illuminance up to a limiting value Results are presented for one observer with a Landolt C as a target and for two observers with a grating as a target The curves are quite different for the two different targets Acuity begins to rise rapidly at a lower luminance for the grating, but the maximum level reached is higher for the Landolt C These results demonstrate a very important point Many of our indices of visual function, such as visual acuity, are highly dependent upon the specific measuring technique employed. It is incorrect to think of these indices as representative of some invariant quantity which is independent of the procedures employed in its measurement

There is a significant difference between the use of letters and the use of gratings or Landolt rings In the case of the latter two, the nature of the figure is known The task is one of identifying the location of a character of the

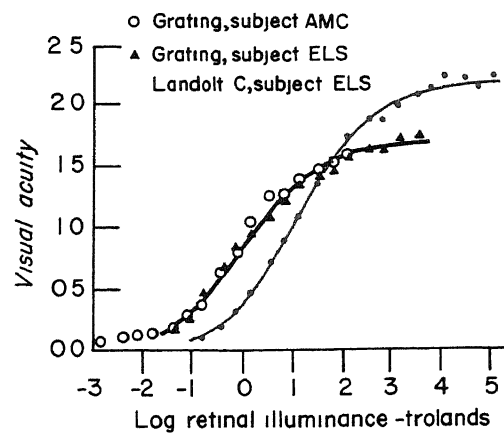


Fig 8 105 Visual acuity as a function of the logarithm of retinal illuminance in trolands for the resolution of a grating test object and the identification of the orientation of a Landolt ring test object Note the more rapid initial rise in acuity with increasing luminance for the grating but the higher final level of acuity for the Landolt ring (Shlaer 1937)

figure which the observer knows to be there at the outset In the case of a letter, the task is one of identification of a given test figure from among a number of possibilities, 26 if the entire alphabet is employed Identification when many possible alternatives exist is somewhat more difficult than the detection of a gap in a ring or orientation of the lines of a grating This is true, at least, when the number of alternatives is relatively large

Form Discrimination The ability of the visual system to aid in the identification of one object out of a number of alternative possibilities is perhaps even more interesting than the fineness of detail in the image on the retina which may be discriminated Obviously, the fineness of visual detail which can be discriminated is an important characteristic of the visual system in relation to its ability to discriminate form and recognize objects, but the way in which the visual system serves in the correct identification of a given letter is quite impressive, even when the size of the image of the letter on the retina is relatively large In this instance, we assume that the basis for discrimination is more dependent on cortical process than on the retinal process

We have already seen that there is not a point-to-point correspondence between individual cells of the retina and individual cells or groups of cells within the brain Rather, the nature of the signals transmitted by the million odd fibers of each optic tract is influenced by interactions between elements within the retina Numbers of receptive cells in the retina in a roughly circular array are included in the receptive field of a given ganglion cell Fibers from ganglion cells terminate on cells of the lateral geniculate body and the axonal processes from these terminate on cells of the cortex Many of the radiations from the geniculate body may terminate on a single cortical cell and there is evidence that the receptive fields in the retina associated with many of these radiation fibers are arrayed along straight lines There are thus cells within the cortex which respond essentially to line elements of stimulation within the retina It is possible that the discrimination of form, at least in its early cortical stages, depends upon the excitation of sets of such line element detector cells The matter is still one for speculation, but, as we have seen, there is evidence that connections to line element detector cells of the cortex which may exist in the newborn are somehow lost if one eye of a newborn kitten is occluded at birth and remains unstimulated by spatially varying patterns of light until beyond the 3rd month of life Along with this neurophysiological change, there is a degeneration of the visual process such that complex form can no longer ever be discriminated by the eye which was covered during the first 3 months of life

The discovery of cortical cells which respond to line elements of stimulation has prompted the study of a variety of effects which result from prolonged exposure of the human eye to line patterns of stimulation The threshold for detection of a grating pattern is influenced by pre-exposure to a grating pattern with its lines oriented in the same way If an observer looks fixedly at a vertical line pattern, it will have no effect on the subsequent threshold for detection of a horizontal line pattern but may elevate the luminance required for detection of a subsequently presented vertical line pattern or decrease the probability of correct identification of such a pattern at a given luminance level (Campbell and Kulikowski, 1966, Gilinsky, 1968) The effect can be observed with adaptation of one eye and testing of the other, suggesting that the site is in the cortex (Blakemore and Campbell, 1969, Gilinsky and Doherty, 1969)

Spatial Modulation Transfer Function. It has been found useful to specify the characteristics of optical systems in terms of a spatial modulation transfer function The modulation transfer function provides a statement of the range of sinusoidal frequencies which can be transmitted by the system and the extent to which amplitude at any given frequency is attenuated by the system A spatial modulation transfer function deals with sinusoidal waveforms distributed in space rather than in time Thus, a pattern of vertical bands which represent a sinusoidal variation in brightness from left to right at a given frequency and with a given amplitude of brightness difference between maximum and minimum values may be imaged by an optical system on some remote surface Measurements can then be made of the extent to which the amplitude of the brightness is reduced as a result of using the optical system Obviously, if the spatial frequency is very high, i e , the lines of the pattern are very fine and closely compressed, the

optical system may be unable to resolve the lines at all and the image will consist of homogeneous illumination At the other extreme, if the spatial frequency is too low, limitations on the spatial extent of the optical system or of the minimum rate of change in brightness over space which can be detected may prevent the transmission of any information about the low frequency change in brightness For systems where it is legitimate to consider complex spatial detail as composed of simple sinusoidal components in a given spatial dimension, such a specification may be very useful It will permit the prediction of the precision with which a given spatial distribution of illumination in an object may be rendered in the image by the optical system Systems for which this is true are said to be linear

The technique has been applied to the visual system, including those physiological processes which intervene between the retina and conscious discrimination of visual patterns as well as the optical elements of the eye It is not possible to measure the amplitude of some image of a line pattern in consciousness Therefore an alternative approach has been employed in which amplitude of the brightness difference between maximum and minimum points in the object is adjusted for threshold discrimination of the presence of the line pattern (Schade, 1956, Westheimer, 1960, DePalma and Lowry, 1962) It is assumed that at threshold the visual effect is a constant The greater the amplitude of the brightness in the pattern at which threshold is reached, the less sensitive is the system at that frequency Thus, the reciprocal of the amplitude of the brightness in the pattern which corresponds to threshold provides a measure of the sensitivity of the visual system Several functions which illustrate the reciprocal of amplitude at threshold for a range of spatial frequencies are presented in Figure 8 106 There is reasonable agreement among three of the four functions presented for the high frequency cut-off portion of the curves Unfortunately, with increased deviation from threshold the visual system can no longer legitimately be treated as a linear system and therefore the utility of the functions illustrated in Figure 8 106 is limited The overall luminance level of a pattern is an extremely important variable and the high frequency cut-off is reduced with a decrease in average luminance Reduced luminance is also

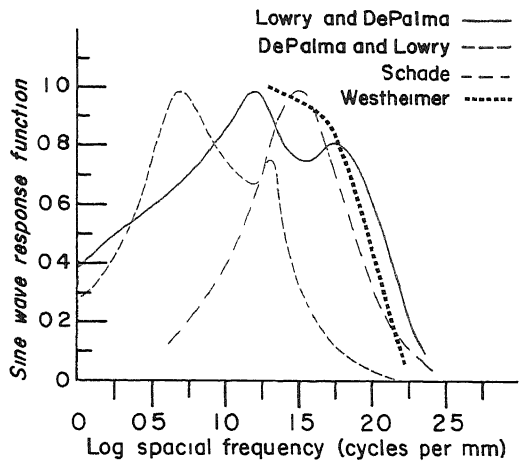


Fig 8 106 Normalized spatial modulation transfer functions for the visual system based on the data of four different experiments Sine-wave response function is related to the reciprocal of the modulation amplitude of the spatially varying sine-wave (Brown and Mueller 1965)

accompanied by adaptation of the eye to a lower level and also by a reduction in the maximum possible physical modulation amplitude of the stimulus

In spite of its limited value, the spatial modulation transfer function of the visual system has given rise to some interesting speculations As mentioned above, recent studies of the adaptation of the eye to line patterns of a given orientation have indicated that there is a selective elevation of threshold for lines of the same orientation subsequently viewed (Campbell and Kulikowski, 1966, Gilinsky, 1968) The effect is to some extent limited to patterns of comparable spatial frequency (Blakemore and Sutton, 1969) Adaptation to a grating pattern will elevate subsequent thresholds for resolution of grating patterns of similar orientations and frequencies which do not differ by more than an octave from the frequency of the adapting pattern (Blakemore and Campbell, 1969). This kind of adaptation is direction-sensitive and shows interocular transfer That is, if the adapting stimulus is presented to one eye, the subsequent effect may be observed with the other eye The directional sensitivity suggests the possibility that line element detectors in the cortex are involved and the interocular transfer capability of the effect provides further support for the conclusion that it is indeed cortical The possibility has been considered that there are specific neural systems for limited spatial fre-

quency ranges (Campbell, Nachmias, and Jukes, 1970)

So far, we have only begun to understand the basic visual functions which underlie such a process as complex as form discrimination It is now clear that aspects of visual information such as line elements and their orientation, motion (Baumgartner et al , 1964) and degree of visual disparity (Barlow et al , 1967) are "encoded" by specific cortical cells That is, certain cells seem to respond specifically and with a high degree of selectivity to such aspects of visual stimulation At the moment, such ideas are purely speculative, but it seems reasonable to suppose that higher order cells or assemblies of cells may become associated with such complex visual information as the identity of a face

Registration of Complex Spatial Information The capacity of the human observer to register information which is received visually seems to be almost beyond measure In a recent series of experiments (Haber, 1970) it was demonstrated that complex photographs numbering in the thousands, which had been seen for only a brief exposure, could later be identified with an accuracy well over 90% correct from among other very similar photographs which had never been seen before This is true even when the photographs seen earlier are projected with a left-to-right reversal of the image Somehow, cortical centers associated with the processing of visual information are able to register and identify an incredibly large number of extremely complex spatial distributions which are detected visually

There is some limited evidence (Luria, 1970) that visual information in the form of printed words may be remembered less efficiently The nature of cortical processing of language is involved with the reading of printed material and difficulties in reading may be associated with dysfunctions of cortical regions which play no role in the perception of visual information that is not associated with language Understanding of these phenomena is still very limited It is important to recognize, however, that there are many complications associated with the visual discrimination of form and visually presented information when we are dealing with situations only a little more complicated than those treated in fundamental studies of the spatial resolution capabilities of the visual system.

TEMPORAL DISCRIMINATION CAPABILITY OF THE VISUAL SYSTEM

Flicker Fusion Frequency. One of the oldest ways of studying temporal resolution capability of the eye has been the technique of presenting a series of flashes to the eye and determining the rate of flash presentation at which the stimulation appears to fuse and is seen as a steady light This is known as the *flicker-fusion frequency* or *critical frequency of fusion* (Brown, 1965b) The results of such a study are presented in Figure 8 107 This shows the critical frequency as a function of the logarithm of retinal illuminance (Hecht and Verrijp, 1933) Except at very high luminances, there is a tendency for the frequency of fusion to increase as retinal illuminance increases That is, the eye is better able to resolve temporal discontinuities with brighter stimuli Although it might seem more "reasonable" for a brighter light to be more persistent, this is obviously not the case We have already learned that the critical duration of the eye is shorter the higher the level of adaptation of the observer

The nearly linear relation over a range of approximately 4 log units between critical fusion frequency and the logarithm of retinal illuminance represents a relation which has been called the "Ferry-Porter Law " It is really not a law in that it does not always apply, but it is a reasonable approximation of the behavior of the visual system for flickering stimuli in a number of circumstances The two other curves in Figure 8 107 show critical fusion frequency for regions of the retina outside

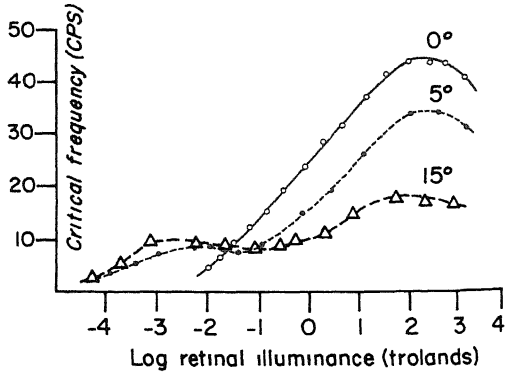


Fig 8 107 The relation between critical fusion frequency and log retinal illuminance for white light for each of three different retinal locations at the fovea and 5 and 15° above the fovea (Hecht and Verrijp 1933)

the fovea The initial branches of these two curves at 15° and 5° eccentricity are attributed to the function of rods The later branches are believed to reflect the function of peripheral cones They are lower by reason of the fact that peripheral cones are sparsely distributed This is presumed to be related to a decrease in their temporal resolution capability The decline of the critical fusion frequency at very high luminances may be attributed to some form of overloading of the retinal receptors Effects of adaptation, retinal location, binocular as opposed to monocular stimulation and spectral distribution of the stimulus all influence the critical fusion frequency There are a number of interactions among these variables but it is impossible to treat them in detail here Several comprehensive reviews of the subject are available (Brown, 1965b, Piéron, 1965)

Temporal Modulation Transfer Function. Just as a modulation transfer function has been investigated in the spatial domain, so has it been studied in the temporal domain of vision Instead of presenting to the eye a rectangular pulse of light which comes on and off abruptly and is interrupted by complete darkness between presentations, a light stimulus may be employed which varies in luminance sinusoidally with time Under these circumstances, it is possible to vary either the frequency of presentation of the stimulus or the amplitude of modulation of the stimulus The average luminance can be kept constant and modulation amplitude may be increased to a maximum of 100% for a sinusoidal stimulus As modulation amplitude is reduced, the stimulus approaches a steady light of unvarying luminance The results of a series of experiments for each of three retinal illuminations are presented in Figure 8 108 At all illuminations it is evident that as frequency is increased, the relative amplitude at which fusion occurs increases rapidly That is, the eye requires a higher amplitude of change in luminance in order to discriminate temporal discontinuity and thus may be said to show a reduction in its sensitivity to temporal variation The left-hand ordinate scale is the reciprocal of relative modulation amplitude At frequencies of over 50 or 60 Hz for the highest illumination employed, it is impossible to discriminate any flicker even at maximum modulation amplitude

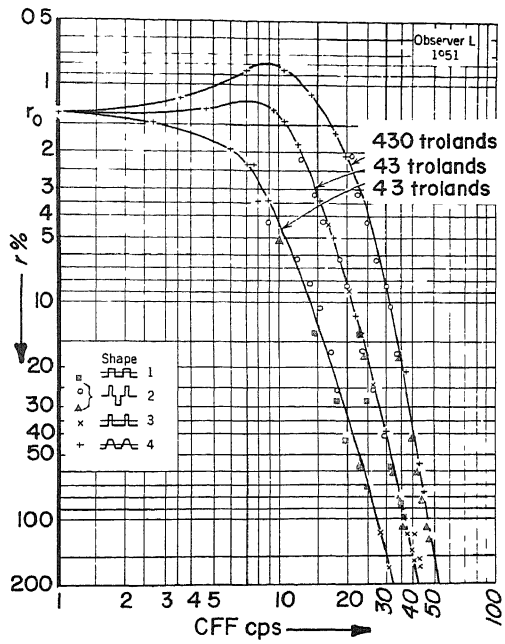


Fig 8 108 Modulation amplitude (r%) of the fundamental sinusoidal component at fusion as a function of frequency for each of four modulation wave forms or shapes The curves represent each of three retinal illuminances as labeled (de Lange 1958)

At the higher retinal illumination, there is the development of a kind of resonance effect A specific frequency appears to provide for maximum sensitivity and at higher and lower frequencies, higher modulation amplitudes are required The peaking or resonance phenomenon is emphasized when larger stimulus fields are employed such that border contrast effects are reduced (Kelly, 1961) For the conditions which obtained when these data were collected, the low frequency modulation amplitudes, expressed in relative terms, are the same for the three levels of retinal illumination investigated

The question must be raised as to the possible practical value of considering the modulation amplitude of a time-varying visual stimulus One very important conclusion which may be drawn is that high frequency variation cannot be discriminated At first, this seems obvious It has been known for a long time that intermittent stimuli are seen to fuse at frequencies which are usually below 100 Hz, with the frequency of fusion depending upon the level of illumination Figure 8 108 provides an additional kind of information, however Four different wave forms of the stimulus light were

employed in gathering these data Their forms are illustrated in the inset of the figure The actual plotted points in the figure represent only the fundamental sinusoidal component of the complex wave forms which were employed For a given average illumination, all of the data points fall on the same curve quite independent of the actual shape of the stimulus wave form The implication of this is that it is only the fundamental sinusoidal component which determines the relation between frequency and modulation amplitude for fusion at a given level of retinal illumination Apparently, the visual system acts as a filter and tends to eliminate the effects of higher order components Although there are special circumstances where this may not be true (Levinson, 1964, 1968), it is true in a sufficiently large number of circumstances to be of some usefulness in predicting the nature of response of the visual system to a time-varying, complex wave form

Perceptual Duration. Techniques other than determination of the time interval in which energy will summate (for the detection of threshold) and the measurement of critical frequency (for the fusion of flicker) have been employed in studying the temporal resolution capability of the eye One procedure uses two stimuli, one visual, and the other either visual or auditory When the second stimulus is an auditory signal, it is usually a click or other sound of very short duration The visual stimulus is presented and the observer's task is to adjust the time of occurrence of the auditory "marker" so that it seems to be synchronous in time with the onset of the visual stimulus In a subsequent series of settings, the observer adjusts the auditory marker so that it appears to correspond in time with the termination of the visual stimulus The difference in time between the termination setting and the setting which marks the beginning of the visual stimulus may be taken as a representation of the perceptual duration of the visual stimulus Studies of this sort indicate that the perception of a visual stimulus is of from 120 to 250 msec in duration (Sperling, 1960, 1967, Neisser, 1967, Haber and Standing 1970, Efron, 1970)

At fairly high levels of light adaptation, a black outlined figure on a white background has a perceptual duration of approximately 250 msec This was ascertained by adjusting the rate at which the black outlined target had to be presented repetitively in order to appear to be on continuously It was actually on for a duration of approximately 20 msec, but observers reported that it was seen continuously so long as it was presented within 250 msec of the preceding presentation on each successive presentation (Haber and Standing, 1969) When the observer is dark adapted and there is no steady background illumination, the "persistence" of a presentation of the same short duration and at the same luminance is increased to over 400 msec

In general, results of experiments of this sort indicate that a light flash of shorter than 250 msec presented to the light-adapted eye will appear to be of nearly 250 msec duration regardless of its duration For the dark-adapted eye, the apparent duration will be somewhat longer For longer light flashes, the apparent duration will conform fairly closely to the actual physical duration It is interesting to note that an interval of approximately 250 msec corresponds to the duration of individual fixations of the eye on a printed page Evidence indicates that an interval of from 200 to 250 msec is required for the processing of information derived from each visual fixation (Haber and Nathanson, 1969) This relatively long time interval of a quarter of a second as contrasted with a critical duration of the order of 0 01 sec in the light-adapted eye is most reasonably attributed to the involvement of cortical processes which must be invoked for the processing of complex information such as letters or words when these are presented sequentially at a fairly rapid rate The temporal resolution characteristics of the visual system are thus quite complex and include a brief limitation at the level of the retina for integration of energy and a more severe limitation which almost certainly involves higher centers in the brain for sequential discrimination of intelligent information Relatively long term temporal summation effects have been demonstrated when the observer's task has been the identification of specific symbols from a larger set of symbols such as letters or numbers (Haber and Hershenson, 1965) The fact that successive presentations of a symbol which shows summation may be presented alternately first to one eye and then to the other

provides clear evidence that such summation takes place at a cortical rather than a retinal level

Discrimination Based on Wavelength

COLOR VISION

The Trichromacy of Color Vision The eye is not nearly as good a discriminator of the wavelength or frequency of light as is the ear of the frequency of sound stimulation Within certain limits, any color sensation may be matched by the mixture of just three primaries if these are appropriately selected spectral lights The only stringent requirement in selecting primaries is that no two of them be complementary That is, that no two can be added together to produce a white or neutral sensation For practical reasons in the matching of colors associated with extreme wavelengths and having relatively high luminance, primaries are usually selected to include a red, a green, and a blue or violet element

Color Mixture. If any monochromatic light is represented in one-half of a bipartite field, it may be matched by the addition of three primary wavelengths in the other half of the bipartite field The relative amounts of three selected primaries required to match any monochromatic lights are illustrated in Figure 8 109 The individual luminances of the specific primaries selected have been adjusted so that when they are mixed in amounts proportional to their relative luminances, they will match a standard reference white

It is clear from Figure 8 109 that a negative amount of one of the three primaries is necessary in order to achieve a match at any wavelength below about 551 nm in the visible spectrum It is, of course, impossible to obtain a negative amount of light under any circumstances The negative values in Figure 8 109 therefore require some clarification In fact, precise matches of monochromatic lights cannot always be achieved by simple addition of all three of the primaries in the opposite side of a bipartite field It is sometimes necessary to add one of the primaries to the wavelength to be matched in order to achieve a perfect match between the two halves of the field

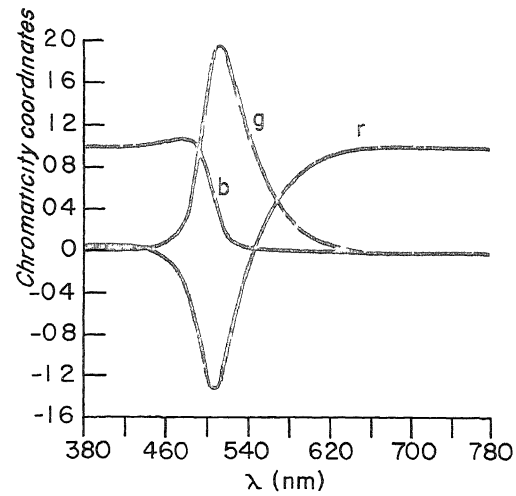


Fig 8 109 Chromaticity co-ordinates of spectrum colors for the standard observer Curves are based on spectrum primaries of 700 nm 546 1 nm and 435 8 nm The reference white is an equal-energy spectrum (Wintringham 1951)

The appearance of the wavelength to be matched is thus altered slightly when a small amount of one of the primaries is added to it It is accepted as a convention that when a given amount of a primary is added to the field to be matched, it is treated as a negative value in the overall addition process

Actually, quite good ranges of colors from violet to deep red may be achieved by simple additive mixture of primaries It is not possible to match all spectral colors without the subtractive technique described above, however

The fact that just three variables need be taken into account for color matching long ago gave rise to the speculation that the visual mechanism must involve three selective discrimination channels in the color vision system (Young, 1807) Such channels might involve each of three different pigments residing in each of three different classes of cones as was very early suggested We have seen above that this type of mechanism probably does, in fact, exist

In its simplest form of operation, such a system might encode information about color in terms of the relative responsiveness of each of the three systems for a given condition of stimulation That is to say, light of long wavelength would stimulate cones bearing red-ab-

sorbing pigments maximally, would stimulate cones with pigments absorbing in the green region of the spectrum slightly, and would hardly stimulate cones with pigments absorbing in the blue region of the spectrum at all The pattern of neural activity in the pathways subserving these three classes of cones would thus be characteristic for a given monochromatic wavelength of stimulation or for a selective distribution of stimulation over a range of wavelengths A theory of color vision based on this kind of reasoning was propounded by Helmholtz (1924) based on the earlier suggestions of Thomas Young (1807)

Opponents Processes. There are certain aspects of the function of the color vision process, however, which belie such a simple mechanism There appear to be inhibitory interactions between certain of the primary color discrimination mechanisms at stages central to the primary receptors When a red light is mixed with a bluish-green light, the combination tends to be achromatic if the components are appropriately selected Similarly, addition of a yellow light and a blue light may yield an achromatic product Individual wavelengths, which when mixed together tend to neutralize each other or produce a white stimulus, are said to be mixture complements It is as if they are stimulating elements within the color vision detecting system which work in opposition and hence tend to cancel out the effects of complementary stimuli when these are presented simultaneously

If the eye is exposed for a period of 5 or 10 sec to a bright red light and this is followed by a white field, the latter will appear greenish in hue If the original red stimulation was smaller in size than the subsequent white field, a green region will be seen against the white field corresponding in size to that of the original red stimulus. Such successive color contrast effects can also be seen dramatically following exposure to a green light which induces a red hue in a white field, or a yellow light which induces blue, or a blue light which induces yellow The precise after-image hue depends upon the wavelength distribution of the initial "inducing" stimulus As reported earlier, electrophysiological processes have been found at the level of the lateral geniculate body which indicate that on certain cells the effect of turning off a red stimulus is the same as the effect of turning on a green stimulus For other cells the reverse may be true Similar effects in individual cells are found for yellow and blue stimulus lights (DeValois, 1965)

If the region of the retina which permits identification of various hues associated with given wavelengths of stimulation is studied carefully, an interesting result is found Red and green may be identified out to a certain distance from the fovea which is very nearly the same for each of these colors On the other hand, wavelengths which typically elicit the response "yellow" or "blue" can be presented at a greater distance from the fovea and will evoke the characteristic color-naming response In a word, the color discrimination ability of the retina for yellow and blue extends out further into the periphery than is the case for red and green These relations are illustrated in Figure 8 110 Finally, if a small patch of color is surrounded by a neutral field and the reflectance of the neutral field is adjusted appropriately, the neutral field will appear tinged with a hue corresponding to the complement of the color patch This is known as simultaneous color contrast (Graham and Brown, 1965) It is attributed to spatial interaction effects which probably reside in the retina, but as yet there is only limited information as to what the nature of these interaction processes may be (Gouras, 1968)

COLOR BLINDNESS

The evidence favoring opponents processes that involve red and green as apparently having opposite effects, and yellow and blue as apparently having opposite effects, are lent further support by the nature of color blindness Although some color blindness is complete, most is not, in that most individuals who suffer from deficiency in color discrimination are able to discriminate some colors although not all Where three primaries are required for the normal observer, many of these color blind individuals require only two They are therefore called *dichromats* as opposed to *trichromats*, the term employed for the normal observer, or *monochromats*, the term applied to the completely color-blind. Dichromacy exists in four specific forms These include *protanopia*, *deuteranopia*, *tritanopia*, and *tetartanopia* (see below) The names applied to these forms of color vision serve only to discriminate them as four different types and have no other scien-

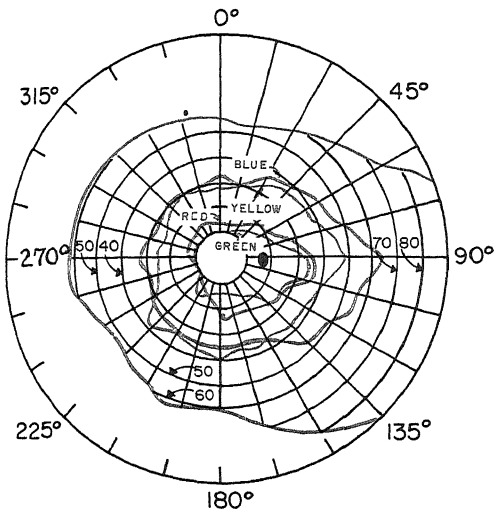


Fig 8 110 Retinal receptive fields in which red green blue yellow and achromatic lights can be detected Note that the zones in which color may be discriminated are smallest for red and green somewhat larger for yellow and blue and largest for the simple detection of light independent of color (Optical Society of America 1953)

tific significance Tritanopes are extremely rare and the status of the fourth type, tetartanopia, is unclear Its frequency of occurrence is sufficiently small that it need not be given serious consideration

Monochromats. There are individuals whose vision is entirely monochromatic, that is, they can match any spectral distribution with a single monochromatic light, provided they are permitted to adjust its luminance Any monochromatic wavelength will serve perfectly well Monochromats, sometimes called achromats, are for the most part found to lack normal cones The vision of these rod monochromats is scotopic and their luminosity function corresponds to that of the dark-adapted peripheral retina of the normal (Sloan, 1954) Visual acuity is low, high luminances are disturbing and there is chronic nystagmus A small percentage of monochromats apparently have one type of cone in addition to rods It has been suggested that these individuals may see in a single color under photopic conditions and in shades of gray under scotopic conditions If this is true, they probably should not be called achromats, although the unitary character of their photopic vision can hardly provide any useful chromatic information Cone-monochromats have been reported with a photopic peak sensitivity at 440 nm (Blackwell and Blackwell, 1957, 1961), suggesting the presence of "blue" cones, and with a peak of 545 nm There is some question as to whether the latter absorption peak represents the function of "green" cones (Pitt, 1944, Weale, 1953)

Dichromats. Both protanopes and deuteranopes tend to confuse reds and greens The colors which they perceive are all in shades of yellow and blue as these are seen by normals This conclusion is based on evidence from cases of monocular color blindness (Judd, 1948, Graham et al , 1961) In such individuals the appearance of various stimuli presented to the dichromatic eye may be matched in the normal eye and thus provide an indication of what the color-blind eye sees relative to normal color vision The color confusions of the tritanope involve yellows and blues Presumably, color blindness involves some deficiency in the mechanisms which contribute to the opponent processes of the normal observer

Extensive studies have been performed on the spectral response functions of color-blind subjects In general, they show relatively little loss in luminosity, and are capable of responding to levels of energy which differ only slightly from those required to stimulate the eye of the normal observer In addition, their visual acuity is almost normal These two facts lead to the conclusion that their deficiency is not the absence of an entire class of cones Rather, it suggests that they may be lacking in one or another of the photopigments found in normal cones

These conclusions have found confirmation in some carefully conducted experiments on spectral response functions (Hsia and Graham, 1957, Wald and Brown, 1965) Spectral response functions representative of protanopes, deuteranopes and normal subjects are presented in Figure 8 111 These functions indicate a systematic loss in sensitivity for the protanope at longer wavelengths It is for this reason that the protanope is often characterized as "red blind" although his actual color perceptions are probably very similar to those of the "green blind" deuteranope The spectral response function for the deuteranope in Figure 8.111 shows a small loss in sensitivity in the short wavelength region of the spectrum, but the loss is not nearly so marked as that for the protanope at the long wavelength end The data in Figure 8 111 were obtained

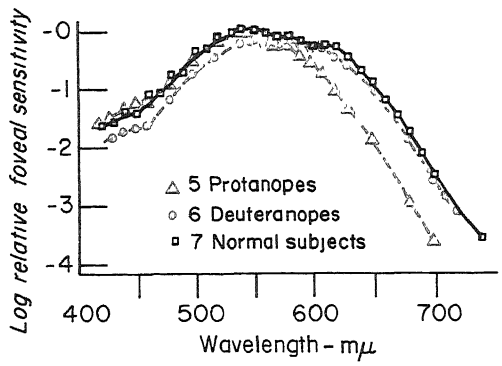


Fig 8 111 Log luminosity for protanopes deuteranopes and normal observers All of the data are expressed relative to the maximum value for the normal observer The protanopes show a luminosity loss in the red and the deuteranopes show a somewhat lesser loss in the green and blue (Hsia and Graham 1957)

without any selective preadaptation They represent the composite spectral response functions which presumably derive from contributions of all available cone pigments as these are combined, probably in the retina

Some indication of the nature of the underlying spectral response function for individual receptors may be obtained by selective bleaching and psychophysical measurements of the spectral sensitivity (Wald and Brown 1965) To isolate the blue receptor, measurements were made against a bright yellow background illumination The green receptor was isolated by making measurements on a background of mixed blue and red, or purple light Measurements of the red receptor were accomplished against a bright blue background The results are presented in Figure 8 112 for a normal observer, a red blind, a green blind and a blue blind observer It is clear that for the normal observer three different spectral response functions are derived against each of the three different conditions of background illumination On the other hand, for the red blind observer, selective adaptation produces only two spectral response functions Against a bright blue background, the function measured for the red blind observer is identical with that measured against a purple background It may be inferred that the red blind observer lacks the long wavelength absorbing pigment Similarly, the green blind observer appears to lack the pigment which absorbs selectively in the middle wavelengths and the blue blind observer seems to lack the pigment which absorbs primarily in the shortest visible wavelengths The color vision of dichromats apparently can be explained in terms of loss of specific cone pigments The selective adaptation studies of Wald provide no evidence for a fusion of two pigment types in a single receptor nor for the convergence of two receptor types on common foci within the system with a resulting loss in their unique individual selectivity

The nearly normal visual acuity of dichromats leads to the conclusion that they are not deplete in total number of retinal cones but that, rather, when a given class of cones is missing it has been replaced by others such that in the protanope there are additional middle wavelength cones and in the deuteranope there are additional long wavelength cones There are insufficient data to charac-

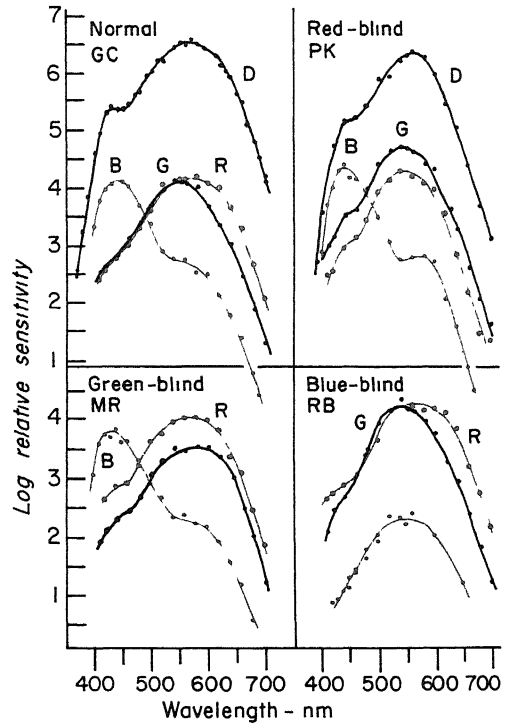


Fig 8 112 Log relative spectral sensitivity functions for normal red blind green blind and blue blind observers Curves labeled D represent the function as measured in the dark-adapted fovea without prior bleaching or selective adaptation Curves labeled *B G* and *R* represent functions obtained against spectrally selected adapting fields which provide for relative isolation of single cone pigments as described in the text It is evident that in the color-blind observers only two spectrally unique functions are obtained (Wald and Brown 1965)

terize the situation in the tritanope It appears that in protanopes and in deuteranopes the mechanisms which stimulate red and green sensations are both stimulated for wavelengths which assume hues of from green through yellow to red in the normal A wavelength of about 490 nm is seen as colorless or neutral by the protanope The neutral point for the deuteranope is at about 500 nm All wavelengths longer than the neutral point evoke a response of yellow in these two classes of dichromats The nature of their color vision deficiency differs in details, but both confuse reds and greens and see colors in shades of blue and yellow Tritanopes see all wavelengths longer than about 575 nm as red and all shorter wavelengths as blue-green, green or sometimes blue

Color Anomaly. There is another class of color deficiency which is termed anomalous trichromacy People with this condition require three primaries to match colors and yet the mixtures which match for them differ in proportions from those made by normal observers There is an anomalous condition for each of the dichromacies, they are therefore categorized as protanomalous, deuteranomalous or tritanomalous In these categories individuals may be found ranging all the way between normal and the dichromat

Hereditary Aspects Color vision deficiency is known to be genetically linked More men are color-blind than women and there are different incidences in different races Among Caucasians approximately 8% of all men are dichromats or anomalous trichromats, only about 0 64% of women are so affected (Hsia and Graham, 1965) Approximately 1% of men and 0 02% of women are protanopes A slightly higher percentage of men, 1% to 2%, are deuteranopes, only 0 01% of women are deuteranopes The occurrence of tritanopes is really too small to be assigned any percentage value Both protanopia and deuteranopia are sex-linked recessive characteristics It is probable that two gene loci are involved on the x-chromosome (Stern, 1960, Francois, 1961, Wald and Brown 1965) It seems reasonable to suppose that the three color vision pigments involve three different opsins which in turn depend upon the operation of three different genes Red blindness and green blindness may represent recessive mutations in which these opsins are not formed

Anomalous trichromia differs genetically from dueteranopia and protanopia The condition breeds true (Waaler, 1927) Wald and Brown (1965) suggest that "the gene that normally specifies the red sensitive pigment is allelic with the recessives for protanopia and protanomaly, and similarly the gene that specifies the opsin of the green sensitive pigment is allelic with the recessives for deuteranopia and deuteranomaly The order of dominance is normal > anomalous trichromia > dichromia"

Tritanopia is apparently not sex-linked A study of 47 tritanopes (Kalmus, 1955) showed 29 to be males and 18 to be females This suggests that the character of the genetic deficiency is autosomal and dominant

Rod-monochromacy is usually said to be a simple autosomal recessive character, i e , not sex-linked In some families, however, evidence for a sex linkage has been found Certain types of cone monochromacy may not be heritable while others probably are (Hsia and Graham, 1965)

Perception of Depth

There is a variety of cues which afford information as to differences in depth, several of which are operative in monocular vision and two of which require binocular vision

Relative Size. The further an object is from the eye, the smaller is its image on the retina Retinal image size thus could afford a cue to depth Nevertheless, retinal image size alone is not a particularly effective cue in the estimation of absolute distance (Gogel, Hartman, and Harker, 1957), even for familiar objects

Linear Perspective. The relative image size of individual objects as a cue to depth may be extended to the image of the entire visual field The texture of objects which recede from the observer in depth grows finer the greater the distance of the particular point under observation Images of parellel lines comprising the edges of three-dimensional rectangular objects tend to converge as they recede from the observer in depth Linear perspective refers to the geometry of projection of a three-dimensional world on a two-dimensional plane and is reasonably considered to be an extension of the relative size factor to the continuum of visual space

Interposition. When two objects are located

at different distances in space, the nearer may intrude between the observer and the more distant object The result will be the partial occlusion of the more distant object and an interruption of its familiar form Ratoosh (1949) expressed the notion formally, stating that a discontinuity in the first derivative of an outline contour affords a cue for discrimination of relative depth of the object forming the contour An illustration of this concept is presented in Figure 8 113 In the part *A* of the figure, one rectangle appears to be located in front of the second in a manner such that it is cutting off the lower left portion of the more distant rectangle The boundaries of the rectangle that appears to be more distant are interrupted by the boundaries of the rectangle that appears closer The contour lines which define what is presumed to be the rear rectangle are thus discontinuous at their point of intersection with the contour lines which define the rectangle which appears to be closer

It is unsatisfactory, however, to attempt to explain interposition entirely in terms of such notions as contour discontinuity A rectangle has four discontinuities, one at each corner, but once the outline is perceived as a rectangle, the discontinuities at the corner have no relevance to depth, they merely confirm rectangularity It is more appropriate to define interposition in terms of the interruption of contours which are expected on the basis of a reasonable assessment of the nature of the figure This is illustrated in part *B* (right) of Figure 8 113 This appears to show a star located in front of a circle Actually, there are discontinuities in the outline contour of the star as well as that of the circle at the point of intersection of the contours of the two figures The discontinuities in the outline of the star conform to expectation on the basis of its shape as a star The discontinuities in the outline of the circle, however, are incompatible with its being a circle unless the star is located in front of it Otherwise, the circle is not a circle but a circular form with a pie-shaped cut-out which just conforms to one of the points of the star For most observers the latter possibility is much less likely than that of a full circle with a star located in front of it (Chapanis and McCleary, 1953) It is thus evident that the cue of interposition is very much dependent upon the observer's knowledge of outline contours of the objects involved

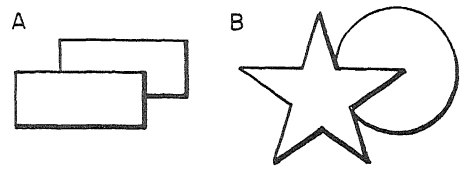


Fig 8 113 Two line drawings which represent the phenomenon of interposition as a cue to the discrimination of relative distance At *A* one rectangle appears to be cutting off a portion of a second rectangle located at a greater distance The contours of the first interrupt the contours of the second rectangle At *B* a star appears to be located in front of a circle In this part of the figure there are discontinuities in the outline boundaries of both star and circle at their intersection The discontinuities in the outline of the star however are characteristic of its shape (Modified from Graham 1965b)

Monocular Movement Parallax. Whenever there is motion of an observer or of objects which are located at different distances with respect to an observer, there is usually an accompanying change in perspective of the objects from the point of view of the observer This is illustrated in Figure 8 114 for a situation where an observer is moving along a line indicated by the arrow Two objects, *a* and *b*, are located in space at the positions indicated The locations of the images of these objects on the retina of the observer are indicated by *a'* and *b'* It is evident that the angle through which the image of *b* sweeps across the retina with motion of the observer from 0_1 to 0_3 is substantially greater than the angle traversed by the image of *a* To the passenger who looks out the side window of a moving car or train, objects near by, such as telephone poles, appear to be moving rapidly in the direction opposite to that of the motion of the vehicle The more distant the object, the less rapid is its apparent motion in the opposite direction Objects at a middle distance fixated by the observer may appear to be stationary, and more distant objects will then appear to be moving in the same direction as the observer Even the slightest motion of the head introduces monocular movement parallax effects which provide important cues to the relative depth of objects in the immediate surroundings Smaller movements are necessary for the discrimination of monocular movement parallax in the horizontal direction than in the vertical (Tschermak-Seysenegg, 1939, Graham et al., 1948)

Accommodation. The accommodative re-

sponse of the eye which alters the physical characteristic of the lens in order to achieve a maximally sharp image on the retina could, logically, serve as a cue to depth If appropriate feedback information is available as to the nature of the motor response necessary to achieve optimum focus of the retinal image, this might provide a cue to the depth of the object viewed When all cues are eliminated other than that of accommodation, however, it is found that this alone does not provide a very accurate basis for the determination of depth (Graham, 1965b)

Convergence. In binocular vision, vergence movements of the eyes result in stimulation of the two foveas by light reflected from the point of regard in the visual world As this point recedes in space, the eyes must diverge As it approaches the observer, the eyes must converge Vergence movements are under the control of the extraocular muscle system and the relative tension in the musculature involved is a possible basis for information as to relative depth of the point of fixation Convergence and accommodation change together in binocular vision With increased convergence, the

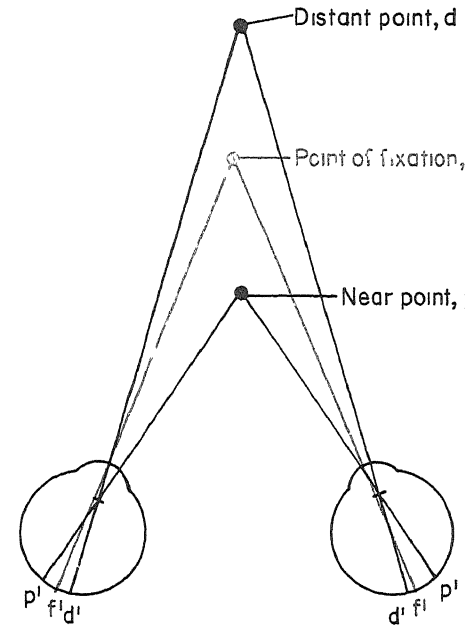


Fig 8 115 Illustration of the doubling of images of objects seen binocularly at depths which differ from the point of fixation For additional information see text

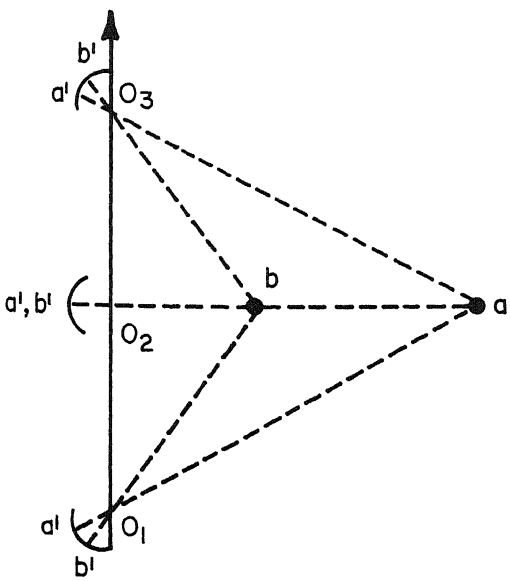


Fig 8 114 Illustration of the relative positions of two objects located at different distances from an observer as he moves along the path represented by the arrow Relative changes in the locations of retinal images of the two objects are shown Such changes in retinal image positions provide the monocular movement parallax cue for distance discrimination

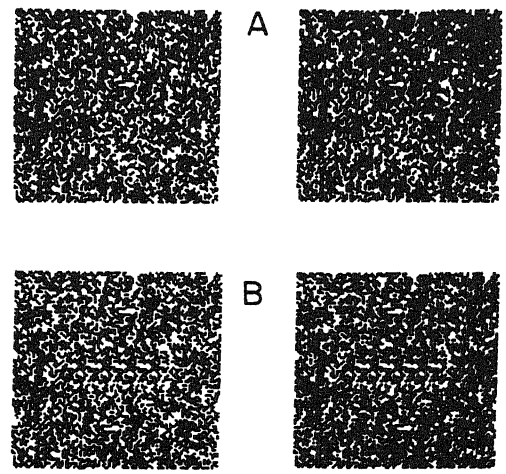


Fig 8 116 Two pairs of random dot matrices which illustrate binocular depth discrimination based on retinal disparity In the upper pair of dot patterns a small square area has been laterally displaced in one of the matrices Otherwise the two patterns are identical The lower two patterns are alike except for three rectangular patterns By first fixating a point closer than the plane of the figure so that there is a doubling of each of the members of a pair of the dot patterns with superimposition of their images on the retina and then carefully adjusting accommodation for clear view of the fused images it is possible to see a square within the pattern which appears to be behind the plane of the page (Julesz, 1964)

eye must accommodate for nearer objects in order to maintain a sharp retinal image A variety of experiments has been performed in which the accommodation and convergence functions are artificially separated The results of these studies do not provide consistent evidence for the comparative evaluation of accommodation and convergence as cues to depth (Graham, 1965b) They do show, however, that convergence like accommodation, is a relatively unimportant cue in the discrimination of depth

Binocular Retinal Disparity The two eyes are displaced in space by approximately 65 mm Thus they provide two different views of the visual world Objects located at different depths are seen in different relative perspectives by the two eyes This can be demonstrated fairly simply as illustrated in Figure 8 115 by fixating the forefinger of the right hand while the forefinger of the left hand is held behind it The more distant finger is seen as double It is clear from the figure that for the left eye, the more distant finger is located to the left of the point of fixation, while for the right eye is is located to the right of the point of fixation The point of fixation is imaged on the foveas of the two eyes The more distant fingertip is imaged on noncorresponding points in the two eyes, to the right of the fovea in the left eye and to the left of the fovea in the right eye (Corresponding points in the two eyes are defined as those points which would be in contact if the two retinas were superposed with the two foveas in contact Thus the temporal retina of the right eye corresponds with the nasal retina of the left eye, etc) If fixation is maintained on the right forefinger and the left forefinger is then brought in closer to the observer, it will again be seen as double, but this time the left eye image will appear to the right of the point of fixation and the right eye image will appear to the left of the point of fixation Again, noncorresponding points of the two retinas are stimulated by the light from the nearer object

In the example given, the effect is a fairly dramatic one in that double images are seen Normally, they are not seen, although with the exception of the object of fixation, parts of the images of most objects in the visual field stimulate noncorresponding points on the two retinas Such retinal image disparity appears to provide a cue as to relative depth When disparity is not great, or when the disparate images are sufficiently removed from the region of the fovea, objects are not seen as double but are seen to be located closer or further away than the point of fixation, depending upon the nature of the disparity Geometrically, there are a number of points in visual space other than the point of fixation which will provide retinal stimulation of corresponding points in the two retinas The locus of these points in space is known as the *geometric horopter* The derivation of the term implies the limiting locus for observation of single images In fact, most points which do not stimulate precisely corresponding retinal locations in the two eyes are not seen as double but are seen as at different depths The geometric horopter was worked out in detail by Johannes Muller (1826) This conception of the horopter is oversimplified, some of the complications which must be considered were worked out by Helmholtz (1924) and by Hering (1861, 1864) A good review of the subject is presented by Ogle (1950)

The recent studies reported by Barlow, Blakemore and Pettigrew (1967) and by Hubel and Wiesel (1970), which demonstrate the existence of cortical cells that respond differentially to certain amounts of disparity particularly in the horizontal dimension, provide a neurophysiological basis for the perception of depth as a result of binocular retinal disparity

Forms can be discriminated in depth solely on the basis of retinal disparity in the absence of any monocular cues as to their identity Julesz (1964) has demonstrated this with the aid of stereo pairs which are comprised of random dot patterns (Fig 8 116) One member of a pair consists of a randomly distributed pattern of dots The other member is identical except that all of the dots within an imaginary outline of some form are transposed laterally with respect to the rest of the dots Dots intruded upon by the transposition are erased The vacated portion of the pattern is refilled with random dots When the two eyes fuse either the dots which have been transposed or the background dots, the resulting retinal disparity of dots not fused is accompanied by the appearance of the form It may appear either in front of or behind the plane of the background depending on the direction of the disparity

5 / TASTE

The sensation of taste is relatively complex Taste is essentially a chemical sense, but the "taste" of food is significantly influenced by responses of pain receptors, and tactual receptors of the mouth, as well as by receptors of smell The way in which a food "tastes" is influenced by its temperature and physical consistency as well as its chemical characteristics One may enhance the taste of food by manipulating it with the tongue while it is in the mouth This serves to distribute the food over different regions of the tongue and mouth in which sensory receptors for taste with differential response characteristics reside In addition, motions of the tongue tend to permit liquids to flow in around the papillae on the tongue, permitting greater access to taste buds in their walls (von Békésy, 1964b) Taste is also influenced by the condition of adaptation of taste receptors Compared with other sense organs, they are highly subject to adaptation, and the taste of any given substance may be significantly influenced by the nature of the materials to which the taste receptors have been exposed in the immediately preceding interval

The presence of certain substances in the blood stream may lead to a sensation of taste peculiar to these specific substances As is true in the case of material taken into the mouth, such tastes tend to adapt out These observations thus indicate that alterations in the chemical constituents of the blood may not only induce tastes but may alter the sensitivity of the individual to the tastes of specific substances The taste receptors themselves are apparently insensitive to mechanical vibration or passive touch, but the texture of foods and the viscosity of liquids certainly influence the nature of the overall experience which we categorize as "taste "

Most persons would agree that taste adds an important dimension to human experience and contributes to our enjoyment of eating or to our rejection of certain foods Although there is some question as to the physiological significance of this for man, tastes may play a highly significant role in maintaining appropriate physiological balance in animals For example, animals with certain food deficiencies have been demonstrated to select preferentially those foods, the taste of which is associated with the substances in which they are deficient (Pfaffmann, 1963, Teitelbaum and Epstein, 1963)

The Sensations of Taste

Concern with the sensory dimensions of taste can be traced back for many years (Boring, 1942) Around the end of the 16th century, nine "basic" tastes were recognized By the middle of the 1700s, this number had grown somewhat by the addition of new tastes and elimination of some of the old A gradual refinement of the list by the elimination of subcategories having components that might be attributed to other than purely chemical stimulation occurred over the years At present, we recognize four fundamental dimensions of taste. salty, sour or acid, sweet, and bitter These four basic sensations were described by Fick in 1864 Acceptance of them has been strengthened by a variety of evidence over the years From these four fundamentals, along with the influences of tactual, pain (Lee, 1954) and olfactory stimulations, the almost numberless varieties of gustatory experiences are derived As we shall see, derived tastes may represent blends of the fundamentals or the effects of interaction which include inhibitory components

Salty tastes are derived primarily from inorganic compounds which ionize in solution The most familiar is sodium chloride, common table salt In addition, the chlorides of potassium, magnesium, ammonium, lithium, zinc, and calcium all taste salty They differ from one another, however, and this is attributed to their stimulation of the receptors for dimensions of taste other than those which mediate

the salty taste In addition to chlorides, the other halides, including bromides, iodides and fluorides, along with sulfates nitrates, carbonates, and tartrates may taste salty The possible contribution of the cation is not clear Compounds of sodium, potassium, lithium, magnesium, and ammonium often taste salty, but sodium acetate does not Heavier halide compounds tend to taste bitter

As is the case for those materials which produce a salty taste, those which produce a sour taste are also electrolytes It is generally agreed that the active agent in this instance is the hydrogen ion The inorganic acids, hydrochloric, sulfuric, and nitric, will taste the same when they are made up in equal concentrations This is not true of organic acids, however Some of these with a lower hydrogen ion concentration than that of hydrochloric acid, for example, may nonethless taste more sour The greater effectiveness of acetic acid than hydrochloric acid for the same hydrogen ion concentration has been attributed to the greater effectiveness of acetic acid in penetrating tissue, and thus, to an actually higher hydrogen ion concentration within the taste buds when acetic acid is employed as a stimulant Another reason suggested by Pfaffmann (1951) is based on the interaction of the saliva with the stimulus Saliva may be a more effective buffer in the case of some stimuli than others Apparently, when acid stimuli are restricted to a region of the tongue from which saliva has been removed, thresholds are directly related to the hydrogen ion concentration

Sweet tastes are primarily associated with organic compounds, specifically the various sugars, along with some polysaccharides, glycerol, and other alcohols, aldehydes, and ketones of the aliphatic series and saccharin, dulcin, and chloroform In addition, some inorganic materials such as lead acetate and diluted alkalis may be sweet to the taste Many of the sweetest tasting organic materials are not electrolytes

As is the case for sweet stimuli, compounds which elicit a bitter taste are primarily organic, particularly the alkaloids (brucine, strychnine, nicotine, quinine, morphine, and others, as well as certain glucosides In addition, bile salts are organic compounds which produce a bitter taste In general, these bitter stimuli are active without dissociation Some inorganic salts including those of magnesium cesium, rubidium, ammonium, and calcium as well as others produce a bitter taste which is attributed to action of the cation in solution Many organic compounds which taste bitter contain NO_2 groups, with two such groups a molecule usually tastes bitter, and with three such groups it invariably does

There appears to be a close chemical relation between many of the stimuli for bitterness and sweetness Derivatives of saccharin, which is intensely sweet, may be quite bitter Phenylthiocarbamide tastes bitter to most individuals and yet it is identical with dulcin, 500 times sweeter than sugar, except for the fact that one oxygen atom of the dulcin molecule has been replaced by a sulfer atom In some cases, isomers of a sweet tasting compound with identical chemical constituents will have a bitter taste

There is a significant relation between the qualitative sensation of taste and the region of the tongue which is stimulated Sensitivity to sweet stimuli is greatest near the tip of the tongue, while sensitivity to bitter stimuli is greatest at the base of the tongue and in the region of the epiglottis Some substances, such as sodium salicylate, rhamnose, and parabrombenzoic-sulfamide give rise to a sweet sensation when applied to the tip of the tongue but elicit a bitter taste when applied to the base of the tongue Magnesium and sodium sulfates cause a salty taste when applied to the tip and a bitter taste when applied to the base of the tongue Apparently the lateral edges of the tongue at about the middle are most sensitive to sour stimuli Salty stimuli appear to be most effective on the tip of the tongue, but elicit a characteristic taste when applied to almost any region of the upper surface and edges, except for the middle region of the top of the tongue, which is apparently insensitive to any chemical taste stimulation These localization phenomena were early interpreted to imply that specific receptors for the four fundamental taste components tend to be, at least to some extent, separated regionally on the tongue The discovery of these geographic distributions of response for sour, salty, bitter, and sweet lends support to the belief that the fundamental components of taste are four in number (Hönig, 1901)

The Peripheral Taste Receptors

ANATOMY AND SENSITIVITY

The primary chemical receptors for taste are the taste buds Most of these are located on the tongue, but some are found on both the hard and soft palate, the mucosa of the epiglottis, and the larynx in the region of the arytenoid cartilages, the mucosa of the pharnyx, and even the mucosa of the lips and cheeks, as well as the bottom of the tongue and the floor of the mouth in some individuals (Bornstein, 1940). In children, taste buds are quite widely distributed and are especially numerous at the tip of the tongue In the adult, the number at the tip is much reduced and the middle portion of the tongue's upper surface is almost devoid of taste buds The taste buds of the tongue are located on the lingual papillae, the many small elevations with which the surface mucosa of the tongue is studded There are three main types of papillae, the filiform, fungiform, and circumvallate The filiform papillae are minute conical structures covering the anterior two-thirds or so of the dorsal surface of the tongue, but they rarely contain taste buds The fungiform papillae are larger, round in shape, and are located mostly at the tip and edges of the tongue Each of these usually holds from 8 to 10 taste buds in the epithelium covering its free surface The circumvallate papillae are the largest of the papillae and are most conspicuous at the posterior portion of the tongue where from 6 to 12 are arranged in a V-formation with the vertex toward the back Their general location is illustrated in Figure 8 117*A* The circumvallate papillae have a central round elevation with perpendicular sides and are surrounded by a sulcus The taste buds are located in mucosa which forms the walls of the circular trench around each papilla There are usually more than 200 taste buds on the perpendicular walls of each papilla and there may be substantially more With advancing age, this number may be drastically reduced (Ary, et al, 1935) A microphotograph of the tip of the tongue which shows various taste papillae is illustrated in Figure 8 118 These have selective responses for the human subject to individual stimulation, both electrically and chemically, as coded in the figure A schematic illustration of a single taste bud in cross section with its microvilli projecting from the lateral wall of a papilla is shown in Figure 8 119 Individual taste buds are oblate, 50 μ in breadth and approximately 70 μ from base to tip They are made up of groups of peripheral supporting cells and elongated receptor cells about 10 μ in diameter The microvilli at the tips of these cells project through the taste pore on the surface of the papilla (Fig 8 120) These are extremely thin, only 0 1 to 0 2 μ in diameter and approximately 2 μ in length (deLorenzo, 1963) The tight packing of the cells suggests that stimulation occurs by reason of contact between the microvilli and solutions which carry taste provoking substances Each taste bud may contain from 5 to 18 sensory cells Innervating nerve fibers penetrate the basement membrane of the receptor organ and arborize upon the surfaces of individual taste

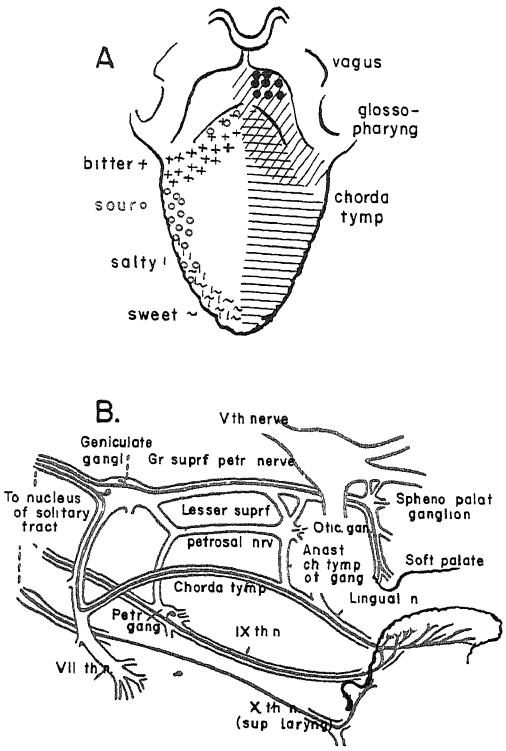


Fig 8 117 *A* Sensitivity of the tongue in different regions for four primary taste qualities Diagonal cross-hatching represents the region innervated by the glossopharyngeal nerve Solid dots show the region innervated by the vagus Horizontal shading represents the region innervated by the chorda tympani (von Békésy 1966)
The pathways of the nerve supplies are shown in greater detail in *B* a sagittal cross-sectional view of the tongue (Modified from Brodal 1959)

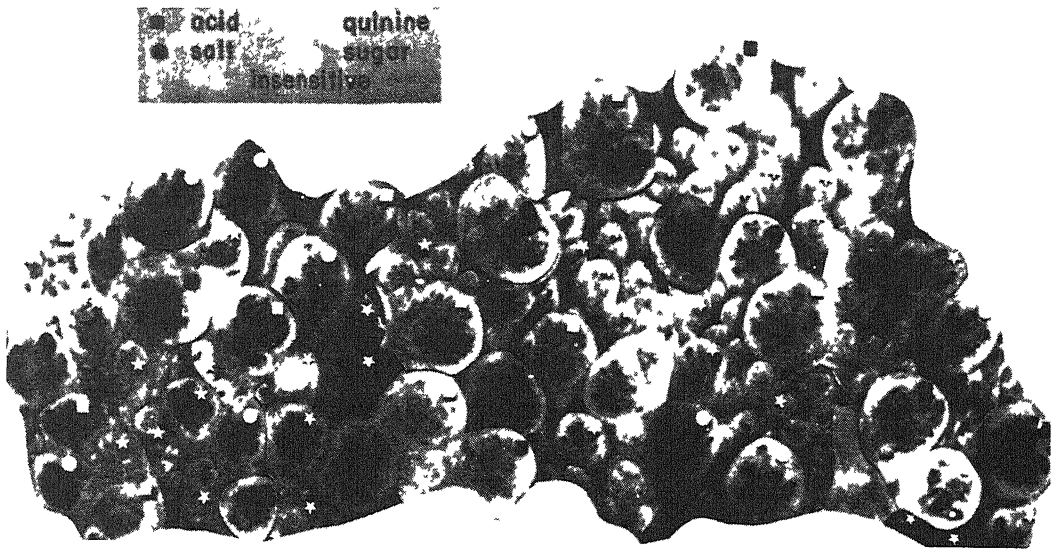


Fig 8 118 Photomicrograph of the surface of the tip of the tongue showing various taste papillae Sensation qualities for electrical and chemical stimulation of each papilla are shown by the various signs Sensation qualities for electrical and chemical stimulation of a given papilla are identical (von Bekésy 1966)

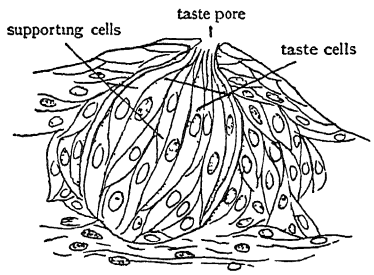


Fig 8 119 Vertical section through a taste bud

cells The suspicion that the sensory cells of taste buds "wear out" and are replaced after a fairly brief life cycle has been confirmed in several experiments (Beidler, et al , 1960, de-Lorenzo, 1963) These studies reveal that the sensory cells arise from division of the surrounding supporting cells and then migrate toward the center of the taste bud prior to degeneration The entire process probably occupies no more than 3 to 5 days for the rat, perhaps 10 to 15 days for the rabbit. Beidler (1963) has suggested that the response characteristics of these cells for a specific stimulus may change during their life cycle, but this raises serious theoretical problems if true (Davis, 1963), and there is some evidence which tends to refute the notion (von Békésy, 1964b, 1966).

The evidence for four fundamental taste sensations has been interpreted to suggest that there may be four distinct types of taste buds, each specialized to respond to certain types of stimuli more effectively than others Although there is some evidence that this may be the case (von Bekésy, 1966), there is other evidence, apparently equally good, which contradicts the idea Kimura and Beidler (1961) recorded from individual taste receptors of the rat and hamster Their results showed that a microelectrode inserted through the taste pore of a single taste bud responds with a graded potential to stimulation provided through a small pipette in the region of the taste pore From a resting level of from −50 to −95 mv relative to a remotely located indifferent electrode, there was an increased positivity by as much as 10 mv upon chemical stimulation Substantial responses could be obtained from a single taste bud to stimuli for salt, bitter, sweet, and sour responses The response might be greater for one chemical modality than another, but variation in concentration could easily equalize the responses Highly specific taste receptors that respond to only one type of stimulation were not found Many of the taste cells did not respond to sucrose but a substantial number did The responses in millivolts of the hamster taste receptor to basic stimuli of four taste qualities are presented for each of eight taste receptors in Table 8 4

Table 8 4
Response in millivolts of the hamster taste receptor to basic stimuli of four taste qualities

Taste Receptor	0 1 M NaCl	0 1 M Sucrose	0 02 M Quinine	0 01 M HCl
1	10 5	2	4 5	—
2	8 5	2	7 5	0
3	9	1	4 5	—
4	11 5	5	8	22 5
5	10	6 5	9	12 5
6	6 5	1 5	5	—
7	9 5	2 5	7	9
8	14	0	5	14

From Kimura and Beidler, 1961

Note that all of these receptors responded to salty and bitter stimulation and that all responded to at least three of the solutions for the four different taste qualities If individual taste buds were strikingly more sensitive in their response to one of the four modalities than the other three, it might be presumed that, even though they respond to a variety of solutions, the individual cells mediate a highly specific qualitative sensory response The data of Table 8 4 however, do not appear to support such a notion These data, of course, are electrical measurements in the peripheral receptors of an infrahuman species It may, therefore, be dangerous to generalize too quickly from these data to theories of taste reception in the human

Several interesting experiments have been performed on the subjective responses of humans to stimulation of individual papillae, both electrically and chemically (von Békésy, 1964b, 1966) von Bekésy (1964b) was able to stimulate individual papillae of human subjects with s small gold-tipped electrode With a tip diameter of $\frac{3}{10}$ mm , only a unitary taste response was obtained Responses of sour, salty, bitter, and one case of sweet were elicited by electrical stimulation of specific papillae Each papilla mediated only one taste quality There was no change in the quality with movement of the stimulating tip over the surface of a single papilla The electrically elicited taste qualities were not identical with those elicited by chemical stimulation This was attributed to the probability that chemical stimuli are effective in the excitation of several papillae and result in a combined excitation with a resulting qualitative change Some papillae were found which appeared to be two simple papillae fused together These mediated dual response qualities Larger papillae were found infrequently which mediated as many as three taste qualities von Békesy (1966) followed up his study of subjective response to electrical stimulation with one of subjective response to chemical stimulation An individual papilla was pulled up with a suction tip and a drop of solution placed on its lateral wall When the tongue was maintained relatively dry in air of 30% humidity, the walls of the papillae apparently would support a drop of test solution without allowing it to flow onto the taste buds of neighboring papillae The results of the chemical stimulation study supported those obtained with electrical stimulation Individual papillae mediated only one qualitative taste response unless they were the larger, apparently compounded, papillae More importantly, individual papillae gave no response whatsoever to solutions which were inappropriate for the stimulation of their particular quality of sensation The quality was not changed by a change in the concentration of solution, and the quality remained the same when a single papilla was stimulated either chemically or electrically These results are very difficult to reconcile with those reported by Beidler and his colleagues as referred to above

There is some additional, albeit indirect, evidence for four differentially responding receptor cells each of which may uniquely mediate a given taste quality von Skramlik (1963) investigated the effects of application of a variety of anesthetics on the tongue He found differences in the change of sensitivity for the different sensory qualities of taste, and differences in recovery time, including in some instances a period of hypersensitivity The bitter "substrate" was most sensitive to anesthesia The effects on the "substrates" which mediated the other three qualities varied with the different anesthesias The results provide a clear indication that the four "substrates" (receptors?) are quite independent Nevertheless, the question of whether there are just four unique receptors for taste in the human gustatory system or whether there are many variations in the taste receptors (Erickson, 1967) must remain an open question at this time

NEURAL INNERVATION OF THE PERIPHERAL RECEPTORS

The nerve fibers which penetrate the basement membrane of taste buds are dendritic branches of neurons, each of which innervates a number of cells They arise in the seventh and ninth cranial nerves The chorda tympani branch of the facial nerve (VII) sends fibers to taste buds which are distributed over the anterior two-thirds of the tongue This distribution is illustrated in Figure 8 117*A* by the horizontal lines on the right hand side of the figure The glossopharyngeal nerve (IX) sends fibers to taste buds which are in the posterior one-third of the tongue This region is illustrated in Figure 8 117*A* by the oblique lines Taste buds which are located on the pharyngeal aspect of the tongue, as well as those on the soft palate, the epiglottis, and in the region of the arytenoid cartilages receive innervation from the vagus nerve (X) This is illustrated in Figure 8 117*A* by the solid dots on the base of the tongue A lateral view of this innervation is shown in Figure 8 117*B* There are individual differences in the nature of the pathway followed by the fibers of the chorda tympani This branch of the facial nerve is so named because it passes over the ossicles of the middle ear not far from the tympanic membrane This location is accessible in animals and has also been accessible in human patients undergoing surgery on the middle ear

The trigeminal nerve (V) mediates sensations of touch, temperature, and pressure from the entire buccal mucosa in addition to the common chemical sense It does not contain fibers which are associated with taste buds, but, as indicated above, aspects of what we commonly call taste are undoubtedly influenced by activity of the trigeminal nerve

A variety of experiments has been performed in which the electrical activity of single fibers of the primary afferent neurons associated with taste buds has been measured In an early study, Pfaffman (1941) showed that various single fibers in the chorda tympani of the cat showed different patterns of excitation by solutions applied to the tongue Certain fibers were most responsive to acid stimuli, others to both acid and salt, others to acid and quinine Sugar proved ineffective in stimulating any of the fibers from which recordings were made The spike discharge patterns of individual fibers did not show any unique characteristics associated with the nature of the stimulus This was interpreted to support the specific nerve energies theory of function of the sensory nervous system That is, the qualitative nature of the information is determined by the specific fibers involved and not by their pattern of discharge Additional work has been done on cats by others (Cohen et al , 1955) as well as on rats, hamsters, and rabbits (Fishman, 1957, Pfaffman, 1955) The dog is of interest because, unlike the cat, it is behaviorally responsive to sweet substances

In a recent experiment, it has been shown that different sugars differ markedly in their effectiveness as stimuli when the criterion of response is the activity in single fibers of the chorda tympani, or the integrated response of the chorda tympani nerve bundle (Andersen et al , 1963) Using 0 5 M solutions of both monosaccharides and disaccharides, it was found that the most potent stimulus was D-fructose followed by sucrose, D-glucose, maltose, D-galactose, and lactose The order for the monosaccharides was D-fructose, D-glucose, and galactose The rank order of potency of the disaccharides appeared to be determined by the potency of that portion of the molecule which combined with the glucose present in all of the disaccharides Some fibers were found which were responsive to only the most potent stimulus, D-fructose The integrated activity of the chorda tympani provided results which were in agreement with those obtained on single fibers as to the relative potency of the stimuli employed These results, in turn, were in good agreement with the results of behavioral studies on human subjects

Some fibers were found which responded both to salt and to sugar Even when fibers were not responsive to salt, prior stimulation with a salt solution tended to inhibit the response to sugar It is tentatively suggested that the inhibitory effect of salt on the response of some cells which respond to sweet stimuli may be a result of the electrostatic binding of salt on the sites which must be occupied by sweet solutes for the stimulation of receptors Among fibers responsive to both sugar and salt, some were more responsive to sugar, others more responsive to salt, and still others equally responsive A third type of fiber was responsive to salt alone A summary of these findings is illustrated in Figure 8 121

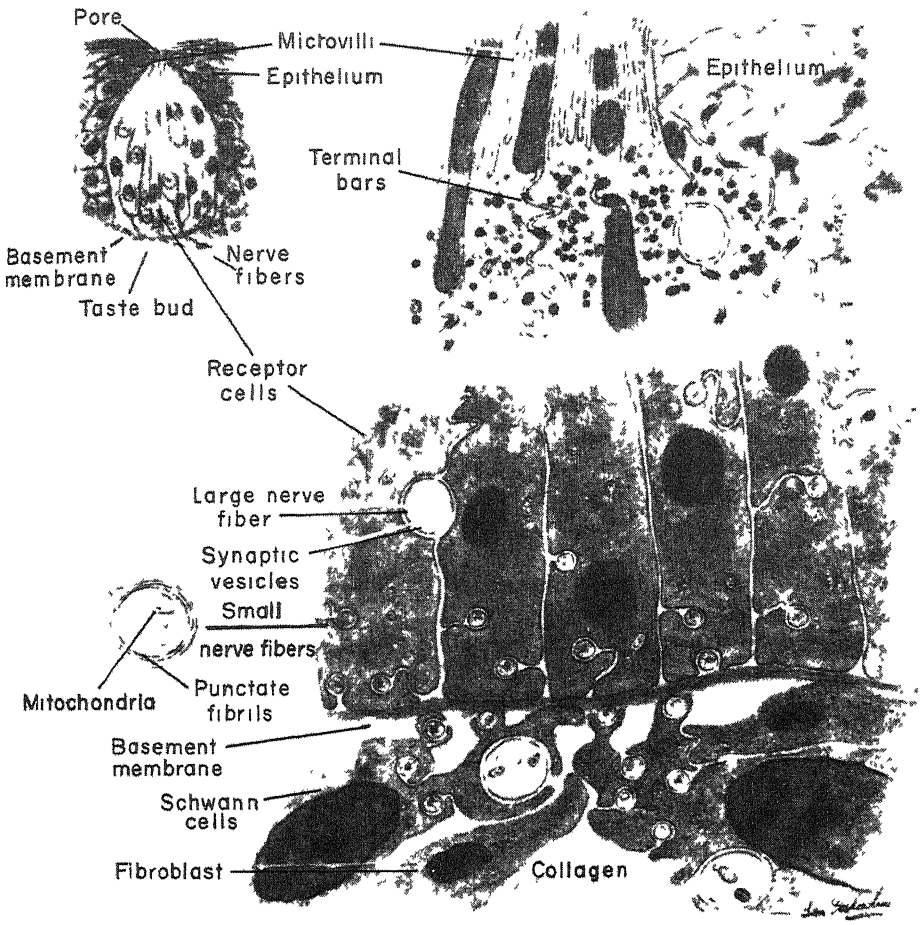


Fig 8 120 Ultrastructure and organization of the taste buds in the papilla foliata of the rabbit (de Lorenzo 1963)

Although some of the single fiber studies indicate exclusive response characteristics associated with a given fundamental taste dimension, it is more typical to find fibers responsive to two or more of the chemical mediators of different fundamental tastes The question arises as to whether any interpretation can be based on these results as to the probable nature of the primary receptors, or their connections to the nerve fibers which carry information from them to the central nervous system It is fairly obvious that many fibers are associated with peripheral terminations on more than one type of receptor, if indeed there are but four receptor types Conflicts between the microelectrode studies of single taste buds and the behavioral studies of von Békésy do not permit any firm conclusion at this time, but some of the single fiber work is suggestive of the possibility that there are but four peripheral receptor types (Ogawa et al , 1968, Frank and Pfaffmann, 1969)

Frank and Pfaffmann have studied the responses of single fibers of the glossopharyngeal nerve and the chorda tympani of the rat in terms of their sensitivity to taste stimuli for each of four different qualities, salt, acid, bitter, and sweet Fibers were classified as responding to a stimulus when they showed at least a 50% increase in response rate during the first 5 sec of stimulation One-quarter of the fibers showed response rate increases of 500% or more to more than one of the stimuli employed An analysis of their data on the glossopharyngeal fibers and unpublished data on chorda tympani fibers has been made Responses occurred in the way they would occur to individual stimuli and each of the six pos-

sible pairs of the four stimuli employed if the sensitivities to the four stimuli are independent and randomly distributed among the innervating fibers Agreement with the hypothesis that sensitivities to the four stimuli are mutually independent events, each with a specifiable probability, tends to confirm the supposition that there are, indeed, just four taste primaries It does not, of course, prove the existence of four independent peripheral receptors

The primary taste fibers in the facial nerve, the glossopharyngeal nerve and the vagus, enter the brainstem and then form a descending tract which terminates in the nucleus of the solitary tract of the medulla Facial and glossopharyngeal fibers terminate in the rostral portion and vagal fibers in the caudal portion (Pfaffman et al , 1961) Secondary neurons which arise in the medulla ascend the brainstem through the mesencephalon in close proximity to other fibers from the face, mouth, and tongue which mediate the cutaneous sense and terminate in the arcuate nucleus of the thalamus The tertiary neurons of the afferent pathway from the thalamus terminate in the cortex near the lower end of the postcentral gyrus in the region of the fissure of Sylvius of the parietal cortex

Physiological studies of these pathways at various locations have been conducted Makous et al (1963) recorded from units of the anterior solitary tract of the medulla and its associated nucleus Responses were found to most of the taste stimuli employed In addition, the same region and immediately overlapping regions produced responses to tactile and temperature stimulation Differences in the response to taste stimuli were observed with respect to the frequency of discharge but no obvious classification of these responses was possible

Higher in the system, single units in the arcuate nucleus of the thalamus may be driven by taste stimulating solutions (Blomquist et al , 1962) Lesions of this region in the monkey thalamus have been shown to produce gross deficits in ability to discriminate differences in taste quality (Patton et al , 1944)

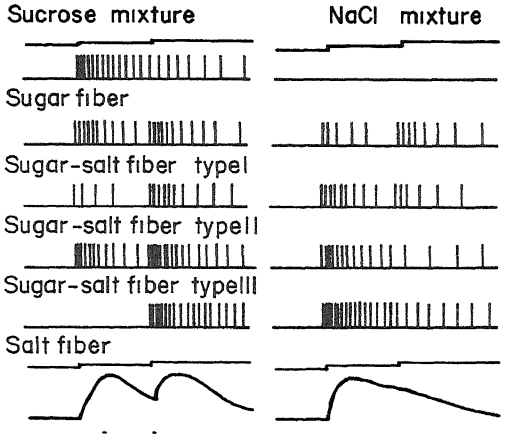


Fig 8 121 Response patterns of various types of fibers which respond to sugar or salt or a combination of the two in solution (Andersen et al 1963)

Experimental Study of Taste Sensation

PSYCHOPHYSICAL STUDIES OF TASTE

Measurements of the absolute threshold concentrations of various taste stimuli are exceedingly difficult to make This is attested by the variability of the results of such studies as these are reported in the literature The methods employed are critical and may greatly influence the results obtained Sapid, or taste provoking, substances may be allowed to flow over an isolated region of the tongue through a glass tube which has a small opening pressed against the tongue The size of the opening and the region of its application will both be relevant variables The temperature of the solution is another variable of considerable importance (Sato, 1963, 1967) Another approach has been to permit the sipping of solutions of material which then can be moved about in the mouth and permitted to stimulate large areas of the surface of the tongue (Anderson, 1950) More precise results may be obtained by the use of a single drop of material (Holway and Hurvich, 1937), or, at the extreme, an effort may be made to stimulate a single papilla Kiesow (1898) attempted to accomplish this with the use of a pointed brush which held the stimulus solution Much more recently, the procedure has been employed by von Békésy (1966) in a study where individual papillae were lifted up from the substrate with a suction tip, and a minute quantity of a chemical stimulus was applied to selected sites on the surface of a given papilla

Absolute thresholds as these were reported

Table 8 5
Absolute Thresholds

Substance	Per Cent Concentration (approx)	Molar Concentration
Sucrose	7×10^{-1}	2×10^{-2}
Sodium chloride	2×10^{-1}	$3\,5 \times 10^{-2}$
Hydrochloric acid	7×10^{-3}	2×10^{-3}
Crystallose (sodium salt of saccharin)	5×10^{-4}	2×10^{-5}
Quinine sulphate	3×10^{-5}	4×10^{-7}

Stevens, 1951

by Pfaffman (1951) are presented in Table 8 5 Values are given for both percentage concentration and molar concentration Since the latter affords an indication of the relative number of molecules of the solute involved, it is perhaps a more meaningful specification Reasonable agreement has been obtained with these threshold values for sucrose and for sodium chloride but widely different results are reported by others for hydrochloric acid, saccharin, and quinine sulfate It is reasonable to assume that the methods of testing, temperature and individual differences all have played a role in these discrepancies which are found in the literature Even the values obtained for one subject differ appreciably from time to time, a phenomenon which may reflect changing concentrations of materials in the blood stream as mentioned above

For some substances the interindividual variability in sensitivity is extremely great A dramatic example is that of phenylthiocarbamide Relative sensitivity to this material is believed to be genetically linked The normal threshold is of the order of from 0 001 to 0 0001% solution For many years it was believed that some individuals were simply incapable of tasting this substance Setterfield et al , (1936) have demonstrated that, when the concentration is sufficiently high, a threshold value is reached for the so-called "nontasters "

The discrimination threshold, or amount of increase in concentration necessary for discrimination of a substance against a background level to which the tongue has been adapted, has been studied As the concentration of the background level is increased the relative amount of increase decreases, as is so for most sensory modalities The ratio between the increment and the background is of the order of one-half to approximately one-tenth The value depends, in addition to the concentration of the background, on the particular taste quality under investigation (Kopera, 1931, Holway and Hurvich, 1937)

Discrimination studies are complicated by the very rapid rate of adaptation of the taste mechanism The weaker the concentration of a substance, the more rapidly its taste will adapt out Even with quite high concentrations, however, adaptation will be virtually complete within 2 to 5 min (Krakauer and Dallenbach, 1937, Abrahams et al , 1937)

For most subjects and for most taste substances, there appears to be an optimum temperature for maximum sensitivity in the neighborhood of 35° C for the human (Fig 8 122) (Sato, 1963) An increase in temperature may result in lower thresholds and increased time required for adaptation (Hahn, 1936) This effect was noted for temperature increases below the optimum level of about 35° C however

It is of interest to consider the way in which stimuli for the so-called fundamental tastes combine to produce complex tastes The richness of our taste experience suggests that most familiar tastes are in fact the result of excitation of more than one type of sensory receptor Nor are the interactions among receptors simply additive Depending upon relative concentrations, the application of a sugar solution to one region of the tongue may enhance the sensitivity to salt stimulation in another region of the tongue or inhibit it A low concentration

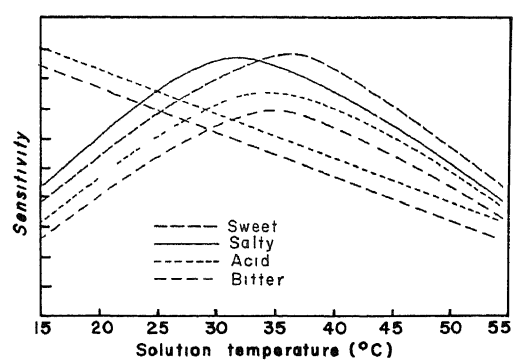


Fig 8 122 Relationship between taste sensitivity of human subjects and temperature of taste solutions Curves represent results for 27 Japanese girls Most showed maximum sensitivity to salt and sugar between 30° and 40° C Half showed a maximum for their sensitivity to acid and bitter and the remaining half showed a regularly declining sensitivity to acid and bitter with increasing temperature (Sato, 1963)

of sugar results in enhancement while a high concentration results in inhibition (Bujas, 1934) Whenever relatively large regions of the tongue are stimulated, interaction effects may be expected no matter what the stimulus Nonetheless, it is possible that some stimuli do afford a taste sensation which is of a fundamental character, that is, basic to the specific receptors for one of the four fundamental taste modalities von Békesy (1966) has suggested that hydrochloric acid may be such a stimulus for "sour," and that quinine may be such a stimulus for "bitter " The subtle differences in taste for different forms of sugars suggest that none of these may elicit the fundamental taste In any case, it might be difficult to distinguish which sweet taste is most nearly representative of the fundamental One device which has been employed by von Bekésy to answer the question is to adapt the tongue strongly with solutions of salty, sour, and bitter substances before applying sweet solutions This procedure results in a more uniform sensory effect of the sweet stimulation which is somewhat different from that elicited by any of the sweet tasting substances without preadaptation, and may provide a method of experiencing the fundamental quality of "sweetness "

Differences in the studies of the response of individual papillae to chemical stimulation by Kiesow (1898) and von Bekesy (1966) may be explainable in terms of a difference in methodology von Békésy was very careful to keep the tongue's surface relatively dry He lifted the papillae to be stimulated out of the surface of the tongue with a suction tip before applying the chemical solution to their sides He observed that the solution remained in place after application without flowing out and providing possible contact with other papillae He explains this on the basis of the argument that the tissue is relatively viscous at its surface and impedes the flow of liquid Were the surface moist, a drop of solution would flow out rapidly and come in contact with other receptors This may have occurred in Kiesow's work Where von Békésy found individual papillae to be, in general, responsive to only one set of stimuli and productive of only one taste quality experience, Kiesow found such exclusively responsive papillae to be in the minority None was found which responded exclusively to bitter stimuli The work of von Békésy is the only work which does report exclusive sensitivities

ELECTRICAL STIMULATION OF THE TASTE RECEPTORS

Electrical stimulation has long been known to be effective in the production of taste sensation With DC stimulation and the anode placed on the tongue, the taste is alkaline This result is attributed to electrolysis and the change in concentration of cations and anions in the region of the electrode when its polarity is reversed An alternating current has been employed in an effort to reduce the electrolytic effect and determine whether direct electrical stimulation may elicit taste With relatively large electrodes, there is a clearly discernible change in the quality of the taste sensation provoked with a change in the frequency of stimulation At very low frequencies, a salty taste may be elicited, at intermediate frequencies, the taste will be sour, and at high frequencies, the taste changes to bitter (von Békesy, 1964b) On the other hand, when the diameter of the stimulating electrode is greatly reduced, a unitary taste experience is provoked, quite independent of frequency These relations are illustrated in Figure 8 123

von Békésy (1964b) used short, pulsatile stimuli of 0 5-msec duration in order to minimize electrolytic effect With a small, gold-tipped electrode, it was possible to stimulate individual papillae These proved chemically responsive to only one taste substance, salt, acid, bitter, or sweet, or to nothing No re-

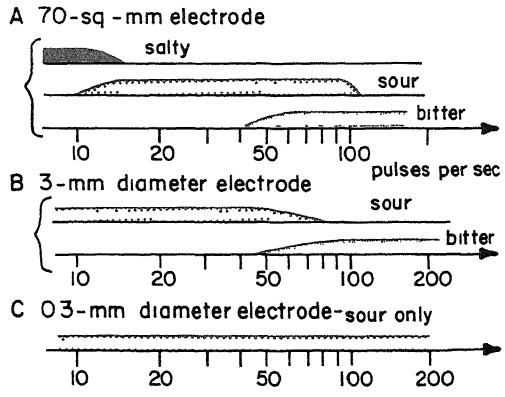


Fig 8 123 Relation of taste quality to frequency of electrical stimulation for each of three stimulating electrodes The quality of taste is seen to change from salty through sour to bitter with a 70mm ^{2} electrode (*A*) the transition is from sour to bitter for a 3-mm diameter electrode (*B*) and taste quality is consistently sour for a 0 3-mm diameter electrode (*C*) With the small elec trode changes in quality could be achieved by selective stimulation of different papillae (von Békésy 1964b)

sponse to more than one stimulus was found except in the case of the double papillae which have been described above These all proved to be combinations of salt and acid sensitivity The response of a given papilla was always consistent and was never found to change with a change in the location of stimulation on its surface Under the conditions of stimulation employed, a taste sensation was evoked only with a positive electrode This suggests that taste induced under a cathode must be attributed to ionization effects With small electrodes and repeated short pulse stimulation, changes in the nature of the taste with increased current did not parallel those changes found with increased chemical concentration On the other hand, with large electrodes employed under conditions where electrolytic effects may be prominent, changes in taste were found to be the same for comparable changes in current or the concentration of a stimulating solution

Using two stimulating electrodes on different papillae, von Békésy was able to show summative effects when papillae of the same character were stimulated, and inhibitory effects for stimulation of papillae representing different taste dimensions, for example, bitter and sweet These results suggest lateral interaction effects which may be either excitatory or inhibitory depending on the nature of the receptors involved The qualitative nature of the tastes elicited by electrical stimulation were comparable for salty, sour, and bitter with tastes produced by appropriate chemical stimuli for these qualities On the other hand, the subjective taste quality for stimulation of a "sweet" papilla differed from the taste elicited by the various sugars It was roughly comparable, however, to the taste elicited by sugar after preadaptation of the tongue with bitter, sour, and salty solutions as described above

Theories of Taste Sensitivity

It is not yet possible to argue with complete conviction for four distinct receptor types which correspond to the four fundamental taste qualities generally recognized Nonetheless, for the human taste process, this seems the most attractive of the various possibilities It is evident that there are interactions among the individual taste receptors within the tongue, and probably also within the other mucosal surfaces in which taste buds are found Taste buds having qualitatively different response characteristics must converge on single primary neurons It is, thus, the pattern of response in the population of neurons which mediate taste which determines the quality of taste (Adrian, 1963) The assignment of a critical role to pattern of response in a population of fibers is almost certainly correct, at least for any but simple tastes, whether one accepts the notion of four fundamental receptors or subscribes to the idea that there may be a variety of combinations of receptors (Erikson, 1965) It is only the two experiments of von Békésy (1964b, 1966) which provide direct support for four unique receptors corresponding to the four fundamental taste qualities, but these experiments appear to have been carefully executed and are relatively convincing They stand in opposition to work which was performed on infrahuman organisms (Kimura and Beidler, 1961) As von Békésy has pointed out (1966), it is reasonable to expect taste to be mediated by a selective chemical action on different organs of taste Certain drugs which impair taste do so selectively Gymnemic acid (Shore, 1892) extinguishes sensitivity to sweet and bitter substances with little if any effect on sensitivity to salty and sour stimuli Cocaine abolishes all taste but the different qualities disappear sequentially in the order of bitter, sweet, salty, and sour These facts may be argued to support the notion of four fundamental taste receptors, although they are clearly not conclusive arguments As indicated above, however, the pattern of responses in individual fibers is also compatible with the notion of four independent peripheral receptors

Like taste, the sense of smell is a chemical sense Perhaps it is even a purer chemical sense than is taste As we have seen, the 'taste" of a food is influenced by tactual stimulation dependent upon the texture and other physical characteristics of food, as well as its odor and its capability for stimulating pain endings which give rise to activity in the trigeminal fibers The contributions of other sensory systems to "smell" are somewhat less This is true for two reasons In the first place, the chemical sense of smell is very much more sensitive than that of taste Secondly, the sense of smell is aroused by chemical material suspended in the air we breathe, and hence, textural qualities are nonexistent It is true that some odors are accompanied by stimulation of the common chemical sensory system mediated by the trigeminal nerve, in which case we may characterize odors as "pungent" or sharp In at least one case, it is probably true that what we believe we are smelling we are in fact tasting This is so for chloroform, which stimulates the taste buds at lower levels of concentration than those required for stimulation of the olfactory receptors Although the conscious concomitants of smell are relatively unspecific compared with the senses of vision and hearing, the sense is an extremely important one for the adaptation of the organism to its environment For many organisms it plays an important role in preservation of both the individual and the species Olfactory stimuli may be paramount as sexual signals for insects and lower animals Although very much more subtle, and probably not requisite for survival, they also play a role in this capacity for the human Information as to the presence of enemies is provided by the olfactory mechanism along with information for the predator which aids him in finding food Food-seeking in many species is partially dependent upon the characteristic odors of appropriate foodstuffs

Olfactory Stimuli

Although it is very difficult to obtain quantitative data on the subject, olfactory stimuli are recognized, for human subjects, to have very strong associative links with past experiences The way in which a characteristic odor may evoke recollection of places and events of the past with remarkable clarity is a fairly common experience

A common problem with the chemical senses is the difficulty of characterizing the stimulus with respect to those aspects of its nature which determine both the quality and the intensity of the sensory experience In general, intensity is related to concentration, but there may be qualitative changes associated with changed concentration as well as intensitive changes In addition, different stimuli, having more or less intense sensory effects cannot easily be categorized in terms of their physical action on the receptors of the system Some substances are particularly effective in stimulating the olfactory sense These include mercaptan and artificial musk which can be detected when present in the amount of one part to several billion parts of the air inspired Other particularly effective stimuli include butyric acid, iodoform, and oil of peppermint It has been estimated that little more than 4×10^{-11} mg of methyl mercaptan/cm ^{3} of air can be detected, and approximately 4×10^{-8} mg /cm ^{3} of artificial musk is sufficient for discrimination of the presence of this material The difference in sensitivity of the sense of taste as compared with that of smell is exemplified by the fact that the concentration of ethanol in solution on the tongue must be nearly 25,000 times as great as the concentration of this material required to stimulate the olfactory sense A relatively small increase in the concentration of alcohol, perhaps by a factor of 3 over that which just stimulates taste, results in a burning sensation by reason

of the stimulation of the receptors for common chemical sensitivity and the resulting trigeminal excitation The concentration of ethanol which can be smelled must be increased nearly 60,000 times before an acrid sensation is elicited in the nostrils

PHYSICAL PROPERTIES OF OLFACTORY STIMULI

In order to be carried to the olfactory receptors, a potential stimulus must be present in air Thus, there is a relation between the volatility of a material or its vapor pressure and its effectiveness as an olfactory stimulus This relation is by no means perfect, however Certain materials, such as musk, have a relatively low volatility, and yet are characterized by a very prominent odor Apparently, this results from a very high effectiveness of the molecule of the substance as a stimulus

Stimulation of the receptor cells in the olfactory mucosa cannot be accomplished unless the stimulants can penetrate the aqueous barrier presented by mucus Effective stimuli are, thus, also water soluble In order to penetrate receptor cells themselves, they must also be lipoid or fat soluble Effective stimuli do have high lipoid solubility and it is, therefore, concluded that solubility in the receptor cell substance is a requisite characteristic There are homologous chemical series, such as alcohols, where solubility changes as a function of position in the series The lower alcohols are highly water soluble, where higher alcohols have a lesser water solubility but high lipoid solubility It turns out that those alcohols in the middle of the series with fairly high solubility in both water and lipoid have the strongest odor Examples of compounds having good solubility in both media include chlorobenzol, brombenzol, ether, and citral Comparisons of various materials suggest that lipid solubility may be slightly more important than water solubility in determining the effectiveness of a stimulus

There are other physical characteristics of materials which appear to be related to their effectiveness as stimuli for the olfactory mechanism So far, however, none of these has permitted any clear-cut theoretical interpretation of the nature of the olfactory process For example, it was first noted by Faraday that strongly odorous materials are also characterized by relatively strong absorption in the infrared region of the electromagnetic spectrum The refractive indices of many odorous materials tend to be very close to a value of 1 5 In addition, the Raman shift of odorous substances tends to be confined to a range of 140 to 350 nm

CHEMICAL COMPOSITION OF ODOROUS MATERIALS

There are certain chemical relations among substances which are associated with their effectiveness as odorous stimuli For example, ethyl, propyl, and butyl acetates all have a characteristic acetic odor Amyl acetate, on the other hand, does not Aspects of the characteristic pineapple odor of amyl acetate can be detected in butyl acetate, however, and the acetic odor is relatively weak in butyl acetate In propyl acetate, the pineapple odor is detectable but the acetic odor is very much stronger In ethyl acetate, the acetic odor predominates Thus, although ethyl acetate and amyl acetate smell quite dissimilar, there is a discernible transition from the odor of one to the other through the series of acetates

There are many quite dissimilar chemical compounds which may have very similar odors On the other hand, identical chemical compounds which differ only in the physical conformation of the molecule may have very different odors This has given rise to the speculation that the shape of the molecule may be a determining factor in its ability to elicit a sensation of odor Such an idea was first put forward by Lucretius (Amoore et al , 1964) It has been explored quantitatively in more recent times and is discussed further below

Of the approximately 30 elements which are found in a free state in nature, all are without odor When heated, arsenic evokes the odor characteristic of garlic but it is odorless as found naturally Relative position of an element in the periodic table appears to have some relation to its odor-producing capability in compounds Sulfur, selenium, and tellurium are all in the sixth group of the table and when combined with hydrogen, methyl, or ethyl groups have disagreeable odors The halogens are all in the seventh group of the periodic table These are associated with a variety of odorless compounds The compounds which stimulate the sense of smell tend to be organic,

but this is not an invariable rule To date, no clearly specifiable dimension of a chemical stimulus has been identified which will permit the specification of either its quantitative or qualitative effectiveness as an odorous stimulus

The Dimensions of Odor

Our language is rich in words for the characterization of odor Odors may be pleasant or unpleasant as well as relatively neutral, and in each of these classifications there are wide qualitative variations It has been estimated that man can distinguish literally thousands of different odors (Wright, 1964) An odor which is not unpleasant may, in increased concentration, become unpleasant although some aspect of its nature is recognized as qualitatively the same

There have been many attempts at the classification of odors Such efforts are usually motivated by the hope that they will provide clues as to the nature of the odor discrimination process, perhaps identifying primary responses from which all of the variations in this sensory dimension are compounded One of the first to develop a set of "primary" odor characteristics was Linnaeus in the 18th century His listing was elaborated by Zwaardemaker (1925) who proposed nine different categories The basis of his proposal was purely subjective, but this alone is no reason to dismiss it Subjective evaluations of sensory experience may provide important clues to the physiological bases of sensory processes A later classification proposed by Henning included six primary odor qualities, five of which corresponded to five of those proposed by Zwaardemaker (Henning, 1916) Another modification, the Crocker-Henderson system, reduced the number to four, three of which corresponded to three of Zwaardemaker's primaries and one of which (acid) was new (Geldard, 1953) It would seem more appropriate to include acid as a taste quality than a smell quality More recently, Amoore (1952), on the basis of a count of the frequency with which various chemical compounds are characterized as having a certain odor, arrived at a list of seven primary odor qualities. Of these, five are found in the earlier list of Zwaardemaker These qualities are presented in Table 8 6 It is interesting to note that only in the case of fragrant or flowery is there agreement among all four of the odor classifications None of the classifications provides sufficient numbers of primaries to permit the range of qualitative discriminations of which the olfactory sense is capable, according to Wright (1964) The basis for discrimination of odors remains unknown It is the hope of investigators in this field that patterns may emerge which will reveal relationships among some aspect of the physical stimuli, the neural activity engendered within the olfactory system, and the qualitative description of the odor experience No satisfactory categorization of primary odors exists at the present time

Table 8 6
Comparison of several classifications of the primary qualities of odor

Zwaardemaker	Henning	Crocker-Henderson	Amoore
Ethereal	Fruity		Ethereal
Aromatic	Spicy		Camphor
Fragrant	Flowery	Fragrant	Floral
Ambrosiac			Musky
Alliaceous			
Empyreumatic	Burnt	Burnt	
Hircine		Goaty	
Foul	Foul		Pungent
Nauseous			
	Resinous		Putrid
		Acid	
			Peppermnty

The Olfactory System

PERIPHERAL RECEPTORS

Olfactory receptors are located within the olfactory epithelium high in the nasal cavity The olfactory epithelium is continuous with the nonsensory nasal mucosa, but brownish-yellow in color due to pigmentation of the supporting cells which it contains This sensory surface is approximately 250 mm ^{2} in area within each nostril and lines the olfactory cleft Its location is illustrated in Figure 8 124

The position of the olfactory mucosa is such that it may not be exposed to the molecules carried by inspired air on each inspiration It has been suggested that the molecules of odorous materials, which are usually volatile, may reach the olfactory mucosa by diffusion It seems unlikely that this is a major basis for the mediation of the olfactory response, however Diffusion is certainly of importance,

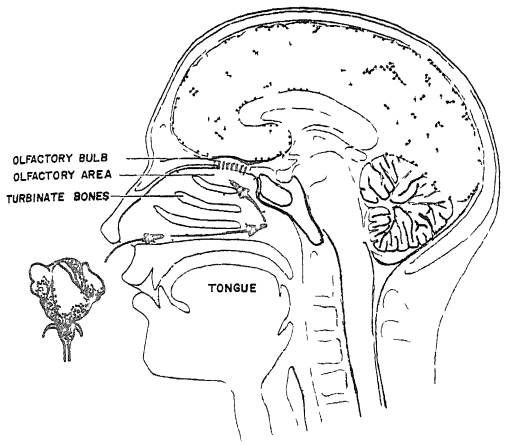


Fig 8 124 Parasagittal cross-section of the head showing the three turbinate bones the olfactory area and the termination of the endings of nerves from the olfactory bulb in the olfactory epithelium The passage of air bearing odorous substances is shown by arrows (Amoore et al 1964)

but maximum exposure of the olfactory epithelium to odorous materials requires an inspiration of more than average forcefulness, or a "sniff" Odorous molecules are carried into the region by eddy currents, primarily on inspiration Stimulation of the olfactory mucosa may also occur on expiration It is possible that some of the contribution of the odor sensing system to the dimensions of taste occurs during expiration when food is being chewed

The receptor cells of the olfactory system which reside in the olfactory epithelium are relatively primitive Unlike many other receptor cells, they serve both as primary receptors and primary neural connections to the next higher level of the nervous system An illustration of the cells of the olfactory mucosa based on electron microscopic examination of this tissue is presented in Figure 8 125 Supporting cells, the sustentacular cells, mingle with the receptor cells The latter may be discriminated from the supporting cells by their longer projection beyond the limiting membrane of the epithelium, and by their relative lack of endoplasmic reticulum The enlarged terminal "rod" of each sensory cell supports between 6 and 12 cilia on its exposed surface in a sepal-like arrangement Microvilli are formed by the plasma membrane, and the cilia of the rods are enmeshed in these The cilia are bathed in a fluid secreted by Bowman's glands In addition to the columnar supporting cells and the terminal rod of the receptor dendrites, there are basal cells in the olfactory epithelium Trigeminal nerve endings have also been identified in this region The cilia can be shown to move, and their motion is enhanced when the mucal substance in which they are surrounded is washed away (Tucker, 1967) The cilia are apparently unnecessary for receptor function, however, and the significance of their number, their orientation, and their movement is unknown In the rabbit at least, it has not been possible to make any clear cut structural differentiations among the olfactory rods (de Lorenzo, 1963) Differentiations have been sought in the hope that they might provide clues to the nature of the discrimination process The only possible basis for classification is the fact that these terminals may contain either a relatively high concentration or a relative paucity of mitochondria

The primary receptor cells are bipolar cells, the axons of which are of relatively small diameter, of the order of 0 2 μ in the rabbit Individual fibers are gathered into small bundles of fascicles which conduct afferent signals upward to the olfactory bulb (de Lorenzo, 1963) Removal of the olfactory bulb in the rat results in complete degeneration of the olfactory neurons (Sen Gupta, 1967)

Electrical Activity of the Olfactory Mucosa. When the olfactory mechanism is appropriately stimulated it is possible to record a negative potential at the olfactory epithelium with a gross electrode (Ottoson, 1963a) This is presumably a reflection of local graded potentials within individual cells of a population of cells which are responding to the stimulation (Ottoson and Shepherd, 1967) This potential was reported by Ottoson in 1956 and is sometimes called the Ottoson potential (Fig 8 126) More commonly, however, it is known as the electro-olfactogram or osmogram (EOG) In as much as the small diameter of the individual fibers of the olfactory receptors made it difficult to record from individual units until very recently, the EOG has provided a useful index of the responsiveness of the olfactory system to stimulation More recently, however, it has been possible to record the pattern of spike activity in single units with microelectrodes (Gesteland et al , 1963) In the frog, unit responses were recorded with metal microelectrodes (Fig 8 127) The resting level of activity was found to be extremely low,

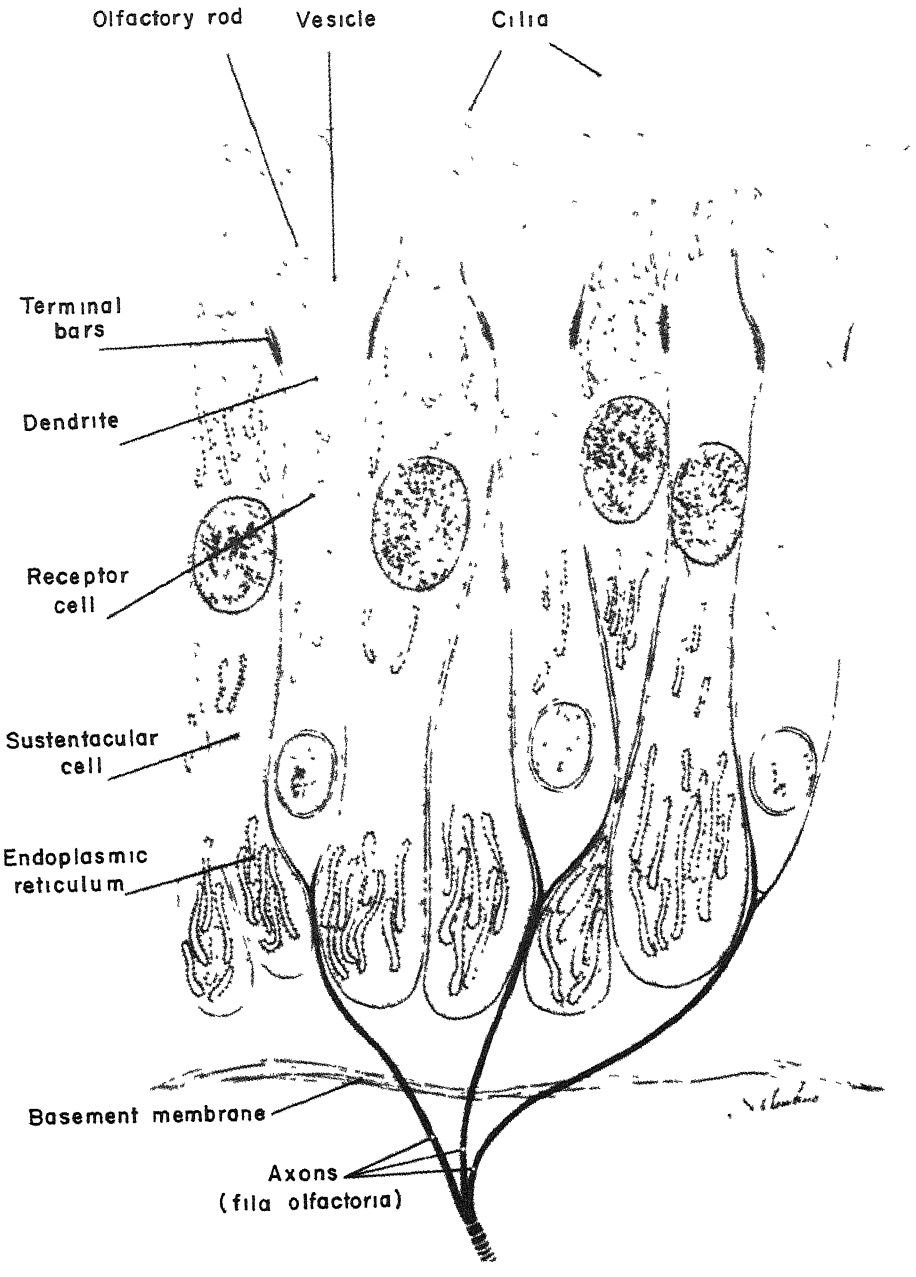


Fig 8 125 Olfactory mucosa showing the relationships of various cell types based on electron microscopic observations (de Lorenzo 1963)

and in many units, there appeared to be no resting level of activity at all On the basis of mode of response to different odorants, eight groups of units were identified from a large number of individual unit recordings These eight groups represented typical response patterns to the different stimuli There was considerable overlapping of response to different stimuli but, nonetheless, the patterns were considered sufficiently distinctive to permit such classification Bursts of spikes which lasted from 1 to 4 sec appeared at rates of up

to 20 spikes/sec Later work (Lettvin and Gesteland, 1965) confirmed the earlier conclusion that individual units within the olfactory epithelium are differentially responsive to different stimuli It appears, however, that the earlier eight classes which were distinguished may be too few In the more recent paper it is suggested that there are at least 10 different kinds of ordering of activity, and probably less than one million kinds of ordering Obviously, the authors are not yet willing to be very specific For any odor employed as a stimulus, it was always possible to find at least one unit stimulated to a high level of activity, and another stimulated hardly at all or possibly even depressed

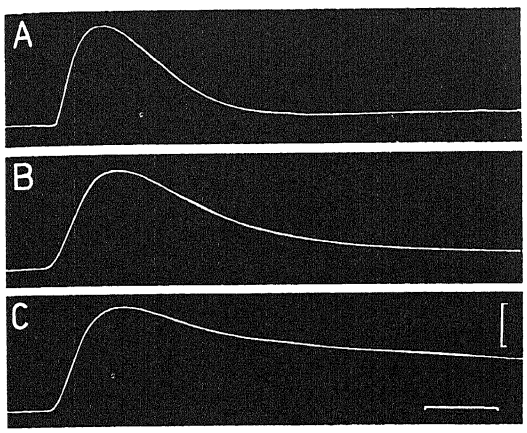


Fig 8 126 Electro-olfactograms (EOG) to different types of stimulation A amyl acetate B butanol C oil of cloves The vertical line represents 1 mv and the time bar represents a duration of 1 sec (Ottoson 1956)

CENTRAL CONNECTIONS IN THE OLFACTORY SYSTEM

Axons of the bipolar cells, whose dendritic terminals are the olfactory rods or primary receptors of the system, cross the lamina propria and join with other axons to form strands of some 20 fibers These ascend in grooves of the ethmoid bone, penetrating the cribriform plate enroute to the olfactory bulb The axonal terminations of these fibers consist of tufted filaments which synapse with the dendritic terminations of cells of the olfactory bulb These terminations are enlarged filamentous bodies known as glomeruli Connections may be with tufted cells, the cell bodies of which are somewhat superficially located, or with mitral cells whose cell bodies are deeper within the structure of the olfactory bulb Granule

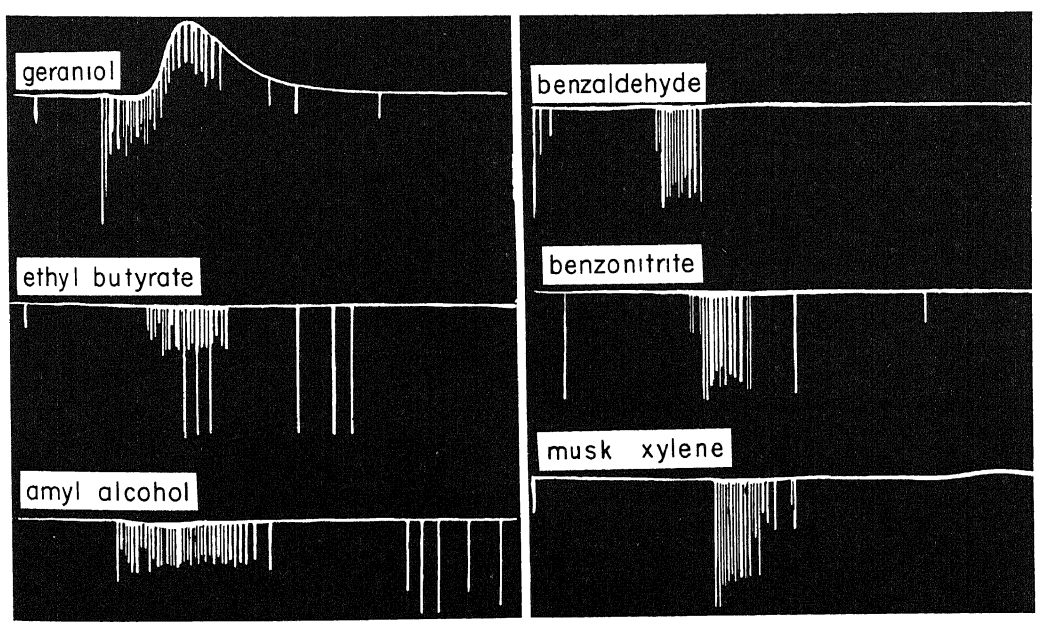


Fig 8 127 Impulse responses to geraniol ethyl butyrate amyl alcohol benzaldehyde benzonitrite and musk xylene All sweep lengths are of 10-sec duration The slow potential is discriminated only for geraniol the other stimuli being too small Several different units may be discriminated on the basis of amplitude of response (Gesteland et al 1963)

cells are located near the center of the olfactory bulb in the core white matter Early microscopic studies suggest that axonal branches from mitral cells reflect back upon the layer of synaptic terminations of the primary cells These cells may be responsible for inhibitory processes which have been observed in the olfactory bulb Connections have been reported between mitral and granule cells in both directions The action of mitral cells on granule cells appears to be excitatory, while the reverse relation results in inhibitory action (Rall et al , 1966) Fibers from tufted cells and from mitral cells project from the olfactory bulb, extensions of some of these fibers form the myelinated fibers of the olfactory tract Section of the olfactory tract results in the degeneration of fibers from mitral cells, but not of those from tufted cells This suggests that the mitral cells alone are the secondary neurons which send projections centrally from the bulb Fibers from the tufted cells may cross to the opposite side of this bilaterally symmetrical organ

Connections within the olfactory bulb as well as cross connections between the two hemispheres are relatively complex There appears to be a plexiform arrangement of fibers arising out of the olfactory epithelium and consequently no point-to-point correspondence between regions of the epithelium and the bulb itself For the rabbit, it has been estimated that there are approximately 1000 primary nerve endings/mitral cell (Rall et al , 1966) von Békésy has suggested that the interactions mediated by lateral interconnections of the two hemispheres provide an olfactory analog of the mechanisms of localization in hearing (1964a)

The electrical activity of the olfactory bulbs has been studied extensively by a number of investigators Adrian (1951) reported some stimulus specificity associated with different regions of the olfactory bulb Heptane and other related chemicals were found more effective in stimulation of the posterior regions of the bulb, while amyl acetate and related substances were more effective in the stimulation of the anterior region Thus, fruity substances are believed to be associated with anterior stimulation and oily substances with posterior stimulation Recordings were made extracellularly with microelectrodes The fact that there is some topological relation between the bulb and the olfactory mucosa, even though it is not a point-to-point relation, suggests a spatial organization of response properties at the level of the olfactory epithelium This work has been confirmed and extended by Mozell and Pfaffmann (1954) and Mozell (1958) The regional specificities were found to be maintained with variation in the concentration of the stimulating substances In more recent investigations of this problem with chronic implants in the rabbit, Moulton (1963, 1965) obtained results which lead to the conclusion that there is spatial and temporal patterning, but he was unable to make any generalization concerning the relative effectiveness of the different odorants employed in the stimulation of the various regions of either the peripheral epithelium or the olfactory bulb

The presence of efferent fibers which appear to regulate the level of response at the olfactory bulb serves to complicate the problem of understanding the olfactory mechanism (Cragg, 1962, Kerr and Hagbarth, 1955) Moulton (1965) was able to obtain much less variable responses from the bulb after sectioning of the anterior olfactory regions through which efferent fibers passed to the bulb Regulatory effects on the olfactory system may also be mediated by extrabulbar efferent pathways There are autonomic fibers which innervate the olfactory mucosa independent of any bulbar connections (Tucker and Beidler, 1956)

Classically, the central projections from the olfactory bulb have been assumed to go to the rhinencephalon, or "smell brain " It is now known that this region of the brain, the limbic region, is concerned to a much greater degree with emotional states and affect than with the processing of olfactory information The myelinated fiber bundle comprised of axons of the mitral cells makes up the olfactory tract which courses caudad under the frontal lobes Unlike any other major sensory pathways there is apparently no stage which passes through the thalamus Connections from the olfactory bulbs appear to be to the prepiriform area, to parts of the amygdaloid complex, and to the olfactory tubercle Contrary to earlier belief, there are no direct connections to the hippocampus or the septal area It also appears that, although they seem to be closely related in sensory experience, there are no significant cortical interrelations of the sensory systems for taste and smell

The olfactory bulb seems to have a promi-

nent role in a sensory discrimination of a nonolfactory nature It has been demonstrated that x-irradiation provides a very effective discriminative stimulus for the control of animal behavior In addition, x-irradiation may give rise to arousal from sleep or changes in the electroencephalogram (EEG) pattern (Kimeldorf and Hunt, 1965) The part of the body which mediates the response to such radiation has been localized in the rat to the olfactory bulb (Hull, et al , 1965, Dinc and Smith, 1966, and Brust-Carmona et al , 1966) Similar results have been found for the monkey (Taylor et al , 1968)

Studies of Olfaction

PSYCHOPHYSICAL EXPERIMENTS

Precise measurements of the stimulus quantity necessary for the induction of an olfactory sensation are quite difficult There is no generalized intensity dimension such as exists in the case of vision and hearing One can only deal with the concentrations of various chemical stimuli, each of which has its own unique accompanying qualitative sensation A variety of techniques has been developed for the measurement of olfactory thresholds One of the first devices used for such purposes was the olfactometer of Zwaardemaker (1925) This consisted of a glass tube inserted into another tube which was porous and impregnated with an odorous substance By varying the extent to which the smaller tube was inserted into the larger one it was possible to vary the length of the larger tube through which the inspired air was drawn before it reached the nostrils This provided a crude control over the concentration of odorous material in the inspired air Zwaardemaker was able to utilize this device for the comparative evaluation of the thresholds of a number of materials

Much more sophisticated devices have been developed since this early technique was employed Elsberg and Levy (1935) utilized a device in which the amount of stimulus-bearing air was controlled The vessel containing the stimulating substance was compressed and a stimulus was presented in a controlled puff to the subject One problem with this procedure is that the sudden application of pressure to the nostrils provides additional sensory stimulation in a nonolfactory dimension It is possible that the active sniffing of a subject is a better procedure Moreover, the control of this procedure from time to time and between individuals is anything but exact Rather complicated olfactometers have been developed in recent years (Johnston, 1953) These provide for the mixture of known concentrations of odorous material with purified air containing controlled amounts of moisture (Fig 8 128) Rotameters permit the precise control of the mixture of various components Subjects may be trained to open the nasal passages so that air is permitted to flow through the nose and down into the mouth continuously A controlled volume presented at a specific flow rate may be employed A normal sniff is probably equivalent to between 50 and 100 cm^3 and some amount in this range may be adopted as a standard

It is extremely important to control environmental conditions, it was demonstrated long ago that prolonged exposure to relatively pure air prior to the measurement of threshold results in a significant lowering of threshold, by 25% on the average for a large number of substances (Komuro, 1921) An elaborate shielded enclosure has been devised for the complete control of the ambient air (Foster et al , 1950) Studies are undertaken in which stimuli may be presented to one nostril or to two nostrils, or different stimuli may be presented on the right and the left side at the same time

As indicated above, the threshold concentrations required for the detection of odors are exceedingly small Even so, the number of molecules in a cubic centimeter at these low concentrations may be rather large Neuhaus (1954) has estimated that at threshold there are 4×10^8 molecules of ethyl mercaptan in

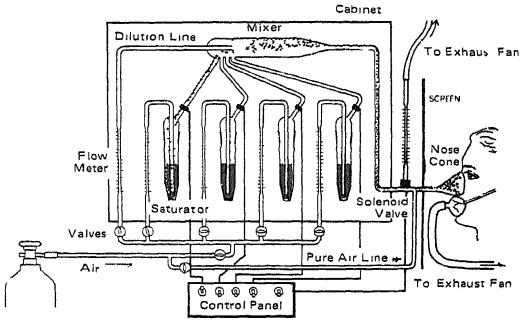


Fig 8 128 An olfactometer which permits the mixture of odors in precise proportions prior to their delivery to a nose cone for sampling Air may be bubbled through liquids to pick up odorous molecules and then diluted with pure air or mixed with air carrying other odors (Amoore et al 1964)

each cubic centimeter of the inspired air It is of interest to consider the number of molecules and receptors which contribute to the detection of odor Computations made by DeVries and Stuiver (1961) suggest that on the order of eight molecules must stimulate each nerve ending, and 40 nerve endings must contribute to the neural activity which is transmitted centrally before an olfactory sensation will arise The large individual differences in structure of the nasal cavity make such computation highly approximate at best

No measure of olfactory sensitivity is any better than the care which has been taken in the purification of materials employed for the purpose Wright (1964) has shown that odors associated with some substances may disappear when the substances are utilized in a highly purified form

The measurement of olfactory thresholds has been shown to bear some relation to tumors with a highly specific location If a tumor is located in the lower region of the frontal lobe, thresholds for odorous substances may be raised and, following exposure to odorous materials, a longer time may be required for recovery of sensitivity than is true for the normal In certain cases of hysteria, with increased intracranial pressure, or in the initial stages of the action of cocaine, olfactory thresholds may be lowered

As is true with all sensory dimensions, an increase in the background level of stimulation is associated with an increase in the amount of the increment which can be detected against the background Direct proportionality of the increment to the background holds over a limited range in many sense modalities and, although only an approximation, is known as the Weber-Fechner "Law " This is also true in olfaction Here, as with other senses, the ratio is relatively higher at low levels and tends to reduce as the background level of stimulation is increased, leveling out at the higher concentrations, and then increasing again The concentration of the increment must be from two times to one and one-third times that of the background to be detected The discrimination sensitivity in olfaction is, thus, not very keen

Adaptation of the olfactory mechanism occurs with prolonged exposure to odor Unlike the visual or the auditory sense, the olfactory sense frequently adapts out entirely That is to say, when it is maintained in a given concentration, a specific odorous material may become completely undetectable after an appropriate length of exposure Complete adaptation to a substance called cumarin in a 0 2% aqueous solution occurs in approximately 2 min Threshold for detection of the chemical following adaptation may be elevated for as long as several hours Adaptation effects may include a central component (Moulton, 1963)

The adaptation is a selective process That is, adaptation to one substance does not necessarily elevate threshold for other substances This suggests that whatever the basis of adaptation, it involves receptors or pathways which are themselves qualitatively selective in function It is obviously not some kind of generalized "fatigue " In some instances, the qualitative aspect of sensation changes with adaptation For example, during the process of adaptation, the odor associated with ionone changes from that of cedarwood to the faint odor of violets, while the odor of nitrobenzol changes from that of bitter almonds to a tarry smell (Geldard, 1953)

One approach to the study of the olfactory process might be to determine very carefully the way in which thresholds for the detection of various substances are elevated by adaptation to other substances This would permit a kind of functional linking of various compounds in terms of their probable action on the same receptor sites within the olfactory system Smelling of camphor to the point of adaptation results in the elevation of threshold for a number of other substances including oil of clove, oil of eucalyptus, and ether, but does not influence the threshold for benzaldehyde Adaptation to benzaldehyde does not influence the threshold for camphor It seems reasonable to assume that these substances are not interacting with the same sensory terminations in the olfactory epithelium

Another aspect of odor which might provide useful clues as to the nature of the olfactory process relates to the way in which different odors combine Such information is almost entirely subjective and must, therefore, be qualified accordingly, but it would seem that there are at least four possibilities as to the method of combination odors characteristic of two substances may fuse into a new and "unified" odor which may have recognizable aspects related to either or both of the components, the odor associated with the combina-

tion of two substances may be recognizable as a blend of the two much as a combination of notes, there may be an alternation of ascendence of one or the other component when two substances are mixed with respect to the prevailing sensation of odor, and one component of a pair may completely mask any response to the other component There has been some question as to whether a fifth possibility may be entertained, that is that two odors in combination may completely neutralize each other According to Zwaardemaker (1925), this is a possibility but the components must be relatively weak in concentration Others have argued against the possibility of any physiological combination which results in neutralization (Henning, 1924)

The highly subjective nature of odors and their tremendous range of qualitative variation limits the way in which they may be studied with animals The obvious approach is one of determination of threshold for various materials, there is a wealth of literature on this particular subject (Wright, 1964) It is clear that many animals are far better endowed than is man for the detection and discrimination of odors For certain substances, such as acetic acid and butyric acid, for example, dogs are capable of detecting concentrations lower than that required by man, by as much as a million to a hundred million times There is also evidence that the qualitative range of discriminable odors is appreciably broader for some animals than it is for man The survival value of the olfactory process is obviously of greater significance to these organisms than in the case of man Actually, the olfactory mechanism of man himself is probably far more sensitive and selective in its discrimination capacity than is necessary in a civilized world There is undoubtedly a great deal more information available to us in the olfactory realm than we utilize or than we need Such information, however, may have been of the highest significance for our forebears

Just as color blindness has provided us with some important clues as to the nature of the color discrimination process, so a comparable condition in olfaction may afford clues as to the nature of olfactory discrimination This is the condition of selective anosmia Some persons show a selective loss for the detection of odors characteristic of specific substances One of these is hydrogen cyanide This has long been associated with some form of genetic determination, although this has recently been called in question (Brown and Robinette, 1967) Selective loss of sensitivity may mean a genetically determined lack of a specific type of receptor or a unique chemical balance which precludes the detection of certain odors or any of a number of other possibilities The systematic study of the condition would appear to be of potential value (Amoore et al , 1968) Guillot (1948a and b) and LeMagnen (1948) have demonstrated that some individuals have specific anosmias for certain musky odors and not for others This is of particular significance in that many investigators have argued for muskiness as one of the primary odor qualities (Amoore, 1952, 1963)

One interesting characteristic of odors is the fact that their strength does not appear necessarily to be related to threshold (Wright, 1964) An odorous substance with an exceedingly low threshold (musk or vanillin, for example) may never have an overwhelmingly strong smell even at very high concentration On the other hand, a substance with a significantly higher threshold, such as ethyl mercaptan, becomes overwhelmingly strong as the concentration is increased A method of studying this phenomenon has been provided by Kruger and his colleagues (1955a) They have developed a scale of odor strength based on varying concentrations of n-heptonal The odor of this substance is quite strong when it is at high concentration It may be diluted with an odorless solvent such that its strength can be adjusted down to a minimal level By selecting appropriate dilutions, a scale of as many values as desired may be created Other odors can then be matched against the standard scale by observers who have been trained in its use Such a procedure permits a kind of quantification of odor strength at levels well above threshold The relative intensities of the aliphatic alcohols for human subjects have been examined with this procedure by Kruger et al (1955b) Their results indicate a tendency for the strength of an odor to decrease as molecular weight exceeds a certain amount As mentioned above, this is associated with a reduction in the volatility of the substance When various members of the aliphatic series were corrected to have equal vapor pressure, it was found that the heavier molecules had slightly higher odor intensities Of course, the method becomes inapplicable

when the extent of dilution of the simpler molecules becomes so great that the dilute solution retains no odor at all

This method may be of particular value in studying the substances with different relations between threshold and strength as concentration is increased Odors with low threshold which do not increase rapidly in strength as concentration is increased are said to be "insistent" odors

Theories of Olfactory Discrimination

There is substantial and convincing evidence that the qualitative differentiation of olfactory sensation depends upon the pattern of response in specific neurons which make up the olfactory nerve and the olfactory tract The way in which encoding takes place at the olfactory epithelium remains a mystery Dimensions of olfactory sensation are complex and have not been shown, conclusively, to bear any clear-cut relation to physical characteristics of the stimulus There does not seem to be anything analogous to the relation between frequency or frequency complex and tone or tonal quality in hearing, or hue and wavelength distribution in vision There is undoubtedly a relation between the intensity dimension of smell and the rate of activity in the neural substrate which mediates this sensation (Adrian and Ludwig, 1938), but the more interesting questions relate to those aspects of the stimulus and those aspects of the receptor mechanism which determine the qualitative nature of an odor

One of the earliest suggestions as to a mechanism which might explain olfactory phenomena was proposed by Lucretius (Amoore, et al , 1964) He suggested that the airborne particles which carry scent have unique shapes which correspond to sites within the olfactory mechanism These were assumed to vary with respect to shape, and hence, to be ideally suited for the trapping of specific kinds of molecules Thus, depending upon the specific sites in which molecules are entrapped, different odorous qualities may be experienced Theories of this sort were surveyed by Moncrieff (1951) Amoore (1952), after a detailed search of the literature on olfaction, catalogued odorous substances according to the language used to describe the odors which they produced He found that a large majority of substances could be included in just seven categories and that each of these had many more substances associated with it than did the other categories which were identified in his survey His supposition was that these most frequently occurring sensory qualities might correspond with anatomical receptor primaries The next stage of his undertaking was to look for some characteristic which those substances having a common capacity for the stimulation of odor might share with respect to their chemical or physical nature His conclusion was that the shape of the molecule was in most instances of a similar form for materials with similar odor-producing capabilities Two of the odor qualities apparently depended upon the electrolytic characteristics of the substance rather than molecular structure Amoore, therefore, devised receptor sites in five unique configurations based on the form of molecules which produce a characteristic odor Early reports of discrimination among neural units of the olfactory mucosa (Gesteland et al , 1963) were interpreted by Amoore in support of his theory (Amoore, et al , 1964) He has evaluated the adequacy of the physical receptor sites he devised, in terms of how nearly they are filled by models of the molecules to which they are presumed to play host, by actual physical test of the extent to which they fill the volume of the theoretical sites The results of this evaluation showed the sites to be rather unsatisfactory in terms of the efficiency with which they could encompass the volume of molecules with similar odor characteristics He has, therefore, abandoned his earlier theoretical site designs (1965), but has shown a definite correlation between the character of an odor and an index of the geometry of the molecule based upon three orthogonal projections of its spatial configuration There appears to be some significance to molecular shape in the determination of odor, but the exact relation of shape to some characteristic of the olfactory receptors is still unknown Amoore's original seven primaries seem to represent an oversimplication of the situation

As mentioned earlier, odorous substances tend to be highly absorbent in the infrared region of the spectrum One theory of olfactory discrimination (Beck and Miles, 1947) held

that local cooling resulting from the absorption of heat by odorous molecules provides a basis for discrimination A serious difficulty with this theory arises out of the fact that certain substances with identical infrared absorption spectra have different odors In addition, attempts to duplicate early experiments which supported the theory have proven unsuccessful (Johnston, 1953) A theory based on the supposition that there is a variety of different enzymes which may be selectively inhibited by different odorous substances was proposed by Kistiakowski in 1950 A small amount of odorous material might block an enzyme reaction and result in the accumulation of a large amount of substance which could provide stimulation of selected nerve endings Objections to the theory have been based on the fact that many optical isomers which should react differently with enzymes provide the same odor (Wright, 1964) It is also argued that too large a number of different receptors would be required for the system and that the temporal response characteristics of the olfactory sense, which are relatively rapid, are incompatible with a primarily chemical basis of detection

The theory might be useful if the olfactory epithelium could somehow accomplish a preliminary discrimination of substances based on their physical or chemical properties This would reduce the large number of different enzyme classes necessary for the range of odorous qualities which can be discriminated Patterning of the distribution of odorous stimuli over the olfactory epithelium may be selective by reason of some sort of chromatographic process (Mozell, 1966) Davies (1965) has argued for variation in the fluid properties of receptor cell membranes as a basis for receptor selectivity Other theories have been proposed based upon the charge-transfer effects between carotinoid pigments in the receptor membrane and specific odor-bearing molecules (Moulton & Beidler, 1967, Rosenberg et al , 1968)

Some years ago (Dyson, 1937), it was proposed that the characteristic resonant frequencies of specific molecules might provide a basis for the determination of the odors which they produced The "Raman effect" provides a method of determining the characteristic vibration frequencies of specific molecules Wright (1964) has argued that vibrational frequencies, expressed in wave number, are not associated with sufficient energy to be of any significance for a mechanism of smell discrimination, unless the wave number is 200 or less Early criticisms of Dyson's theory were based on noncorrespondence of very much higher wave numbers for molecules producing similar odors Recently, Wright and his associates (Wright et al , 1967) have concluded that the frequency components of the activity induced in the olfactory bulb by odorous stimuli are correlated with molecular frequencies of the molecules which constitute these stimuli The results cannot be considered conclusive, however, since the correlation is far from complete when all frequency components are considered

Although there are intimations of progress, the question of how the olfactory mechanism discriminates qualitative differences in odor is far from resolved It is fair to say that different neural pathways and patterns of response among populations of pathways are involved in the qualitative discriminations of odor, but the precise way in which the initial discrimination is made at the primary receptor site is unknown Future research must include a comparative evaluation of electrophysiological studies, psychophysical studies based on threshold and qualitative discriminations for a variety of stimuli alone and in combination, and the individual differences represented by selective anosmias which may be genetic in origin

7 / TOUCH, TEMPERATURE, AND PAIN SENSITIVITY

The successful adaptation of an organism to its environment requires the continuous derivation and processing of a variety of information Although the most richly varied of the sensory dimensions may be vision and hearing, those treated in this chapter are nonetheless important for the maintenance of the health and welfare of the organism Receptors in the skin and other parts of the body which make contact with the external world, including the tongue and the teeth, provide a basis for constantly monitoring the compatibility of the environment

Elaborate mechanisms for the regulation of body temperature are dealt with elsewhere, but without the additional mechanisms which depend upon voluntary actions mediated by conscious response to sensory stimulation, it would be impossible for man to maintain regulation of body temperature over the range of environmental temperatures in which he is able to function The rate of his voluntary activity and the nature of his clothing are important factors which are controlled by conscious processes

Protection of the organism from injury requires information as to the nature of objects and surfaces with which the body may come in contact Such characteristics as hardness and softness and abrasive qualities may not be discernible visually and may require exploratory physical contact The manipulation of objects depends on information as to the nature of the contact made by the body, usually the hands or feet, and changes in pressure associated with changes in the tension of muscles controlling the fingers or perhaps the jaws Feedback systems involving muscle sensors permit highly refined control, but the augmentation of this system by tactual sensitivity is essential for proper function Motor control of the body depends to a significant degree on information based on tactual sensation derived from contact with the surfaces on which the body may be lying, sitting, standing, or walking

Physical contacts may be destructive Deformation of tissue at too high a rate or to too great an extent can cause injury The same is true of objects at a high temperature or, less frequently, at a very low temperature, sources of thermal radiation, and chemical agents Virtually any agent capable of causing tissue damage signals the organism by stimulating receptors for pain There are some exceptions, such as ultraviolet radiation and ionizing radiation It is only in the very recent history of man that ionizing radiation has constituted a significant hazard and it is perhaps for this reason that no "counter measures" have evolved As noted in an earlier chapter, there may be some possibilities for direct detection of ionizing radiation, mediated by the olfactory bulb

The deep pain which is signalled by receptors in the internal organs is qualitatively different and the sensory system which mediates this response serves a different purpose Ideally, its function may be to impose restrictions on activity during a period of recovery from injury In extreme conditions, however, such pain may be completely nonadaptive

In addition to its role in the regulation of behavior and survival, the sense of touch may play a significant role in communication An excellent direct example is the reading of Braille by the blind, a more subtle example is the nature of the communication process involved in the character of a handshake or any physical contact between individuals The sense of touch plays an extremely important role in the communication between individuals which precedes sexual activity in all of the higher organisms

In relatively recent years, considerable effort has been devoted toward the use of the tactual sense as a means of substituting for vision in the case of the blind A large matrix of stimulators may be excited by a system which involves a television camera It has been demonstrated that, with sufficient practice, blind

individuals can do a remarkable job of discriminating the nature of objects in the real world with such a system (Collins and Saunders, 1970)

The Sense of Touch

The sense of touch is primarily a cutaneous sense, although it is usually considered to include the touch receptors within the mouth, particularly on the tip of the tongue, and the receptors located in dermal tissue, tendons, the periosteum, and other deep seated structures which respond to pressure

THE STIMULUS

The sensory endings for touch are known to respond to uneven or nonuniform pressure which results in deformation of the skin or movement of the hairs which grow out of the skin Sensitivity to touch varies in different regions of the body, but may require only very slight deformation or barely perceptible deflection of a single hair on the surface of the body A wisp of cotton drawn over a hairy surface or even over bare skin is sufficient for stimulating the sense of touch A technique for measuring the relative sensitivity of the various regions of the body's surface to touch was developed by von Frey (1919) A series of hairs, varying in diameter, was graded in relation to the pressure required for bending the hair when it is brought in contact with a surface Threshold for touch may be determined in terms of the pressure required to bend a hair Surfaces of the body which are sparsely covered with hair tend to be more sensitive to direct contact Shaving the hair from a hairy surface results in reduction of sensitivity to touch

Minimum pressures required to elicit the sensation of touch vary by a factor of more than 20 to 1, depending on the region of the body considered The nose, the lips, and the tips of the fingers require only 2 or 3 g /mm ^{2} Higher pressures are required on the backs of the fingers, the upper arm, and the inner surface of the thigh Progressively higher pressures are required on the back of the hand, the calf, the shoulder, the abdomen, the front of the leg, the sole of the foot, and the back of the forearm A pressure of nearly 50 g /mm ^{2} is required on the loin It is evident that those portions of the body's surface which are more likely to be involved in contact which will provide information have a lower threshold

RECEPTOR ORGANS FOR TOUCH

In the skin there are some relatively elaborate sensory receptors which undoubtedly play a significant role in touch There are also undifferentiated nerve endings which are involved in tactual response (Fig 8 129) Unfortunately, it has not been possible to find clear-cut functional differences associated with the anatomical differences among the various receptor types Nonspecialized receptors which cannot be distinguished anatomically may mediate qualitatively different responses, i e , pain and

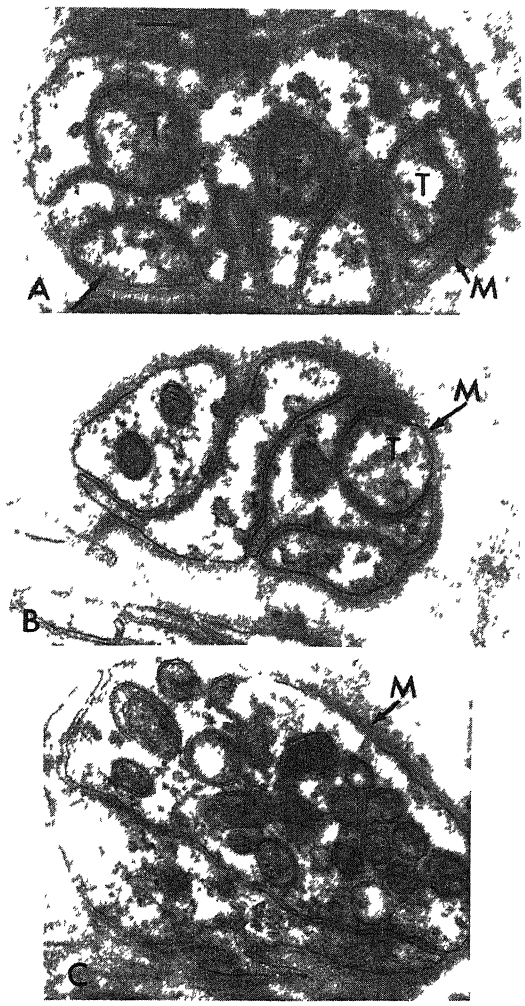


Fig 8 129 Free nerve endings in the tips of dermal papillae of the human digital skin Perineurial sheaths have been lost but Schwann cell coverings and the basement membranes (M) are retained Three terminal branches (T) are shown within one Schwann cell body in *A* A single nerve terminal (T) is shown in relation to a cluster of Schwann cells in *B* An end bulb is shown in *C* with an accumulation of mitochondria varying in size and density (× 51 000) (Cauna 1968)

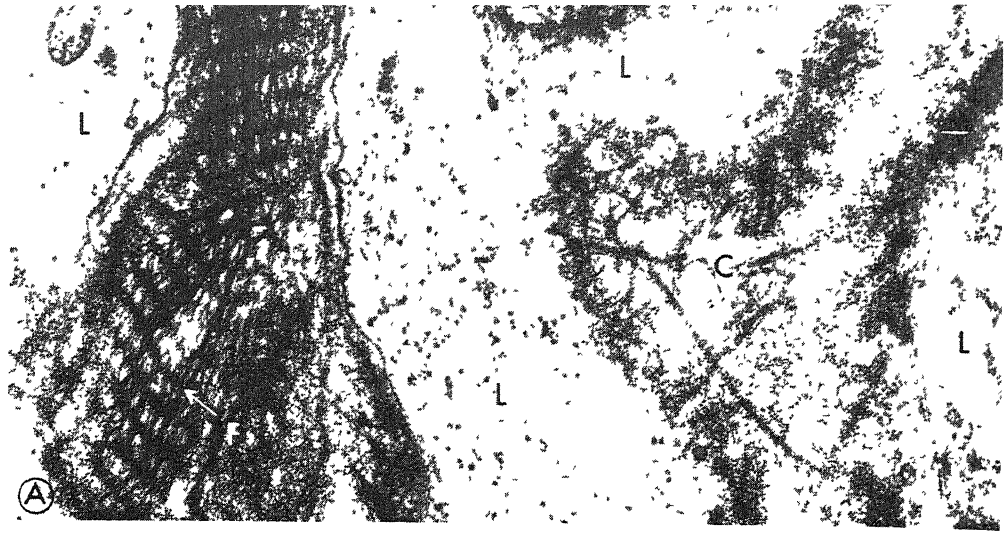
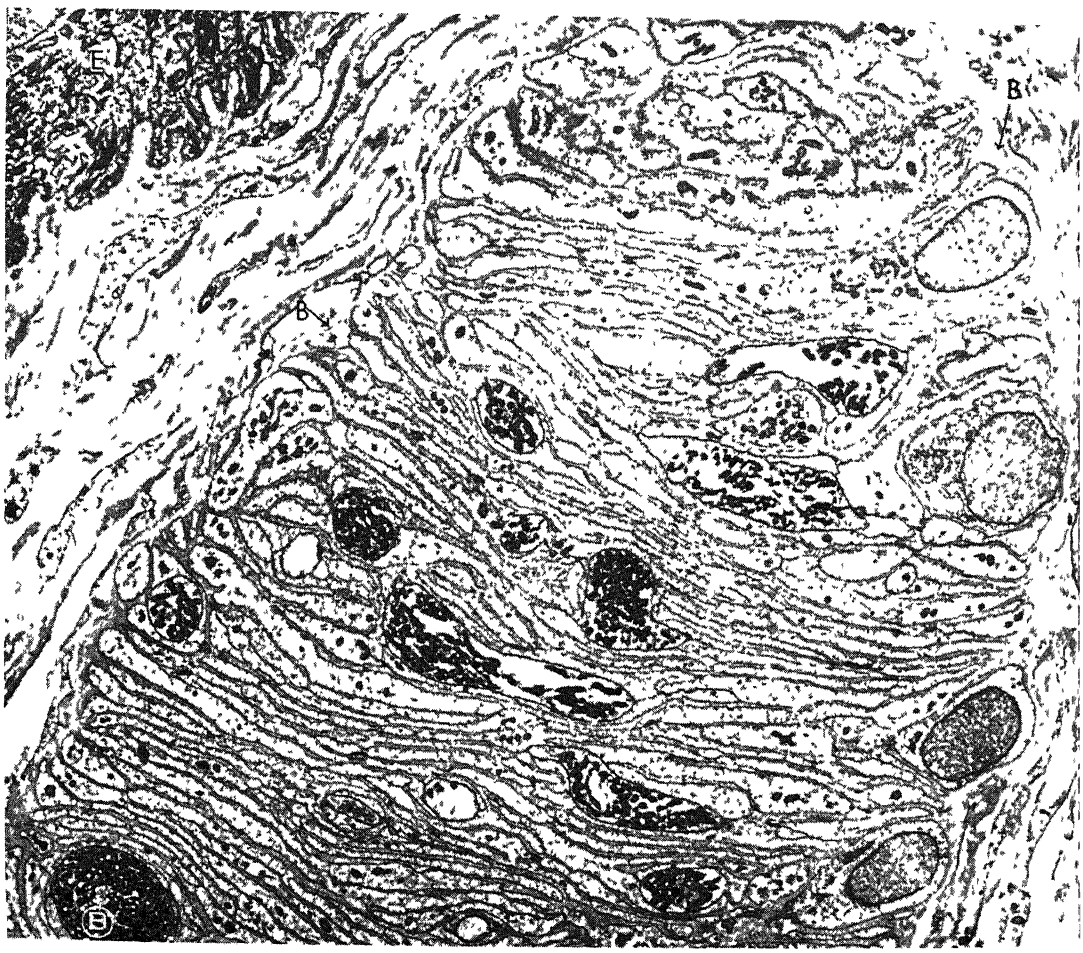
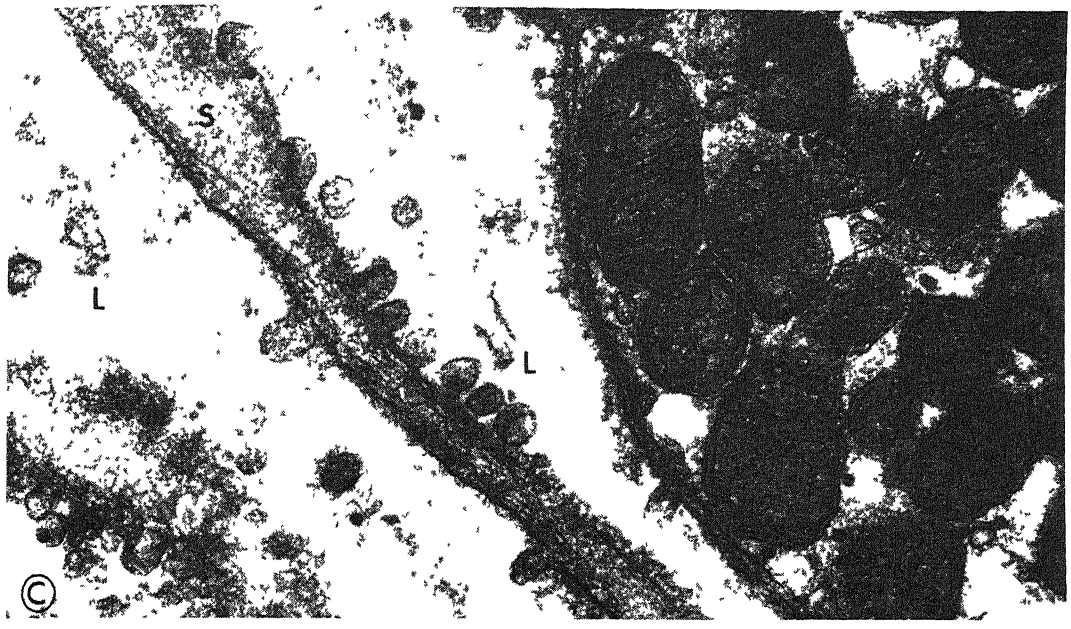


Fig 8 130 *A* oblique cut through a Meissner s corpuscle L laminar cells F bundles of fine fibrils C single collagen fibers of random orientation Human digital skin of a 51-year old male (× 12 500)



B longitudinal section of the superficial part of a Meissner s corpuscle Laminar cells (lighter bands) are separated from one another by an intracellular substance (darker bands) and are interleaved with nerve endings B basement membrane E epidermis Human digital skin from a 21-year old female (× 1500)



C field from a Meissner s corpuscle cut transversely L laminar cells S intracellular substance To the right the nerve ending is shown filled with mitochondria Human digital skin of a 21-year old female (× 32 000) (Cauna 1966)

temperature as well as touch The specialized receptors include Meissner's corpuscles (Fig 8 130) which are found in abundance in the hand, the foot, the nipple, the lips, and the tip of the tongue, Merkel's disks (Fig 8 131) found in the fingertips, the lips, and mouth, and basket-like terminations which surround hair follicles Structures which look like Merkel's disks have been found in epidermal domes (Fig 8 132) which appear to be specialized receptor regions in the skin (Iggo, 1966) Similar structures have been identified in other tissues (Cauna, 1966) Some of these receptors are illustrated schematically in Figure 8 133 There is evidence that the cells which comprise nerve endings change relatively frequently during the life of the individual (Cauna, 1966) This continuing renewal of receptor tissue is also characteristic of the gustatory and olfactory systems

It is obvious from the foregoing that the innervation of the skin of various regions of the body differs with respect to the sense of touch Even in a relatively homogeneous region, such as the back, there are differences. Localized regions, or spots, may be found which are more sensitive to touch or to temperature or to pain Responsiveness to these qualitatively different dimensions of stimulation is not homogeneously distributed



Fig 8 131 Two Merkel s disks in the deepest epithelial cell layer of an Eimer s organ in the tip of the mole snout Merkel s disks are seen directly below the more darkly stained epidermal cells (Quilliam 1966)

Receptive Field. The concept of the *receptive field* was conceived by Sherrington (1906) to describe the selectivity of response of an organism to tactual stimulation. Certain reflexes may be elicited by touch It is not necessary to stimulate one highly specific point, rather, a region may be stimulated and the reflex response will follow For a given reflex, the region in which stimulation will elicit the char-

acteristic response is known as the receptive field for the reflex Sizes of receptive fields vary with the region of the body and with the specific type of reflex investigated The results will differ dependent upon the qualitative nature of the stimulus

Functional Differentiation of Mechanoreceptors Two classes of receptors that respond to touch have been identified The first responds with impulse discharge for a period of only a few hundred milliseconds and then adapts, even though the tactual stimulation is sustained They have a relatively low thresh-

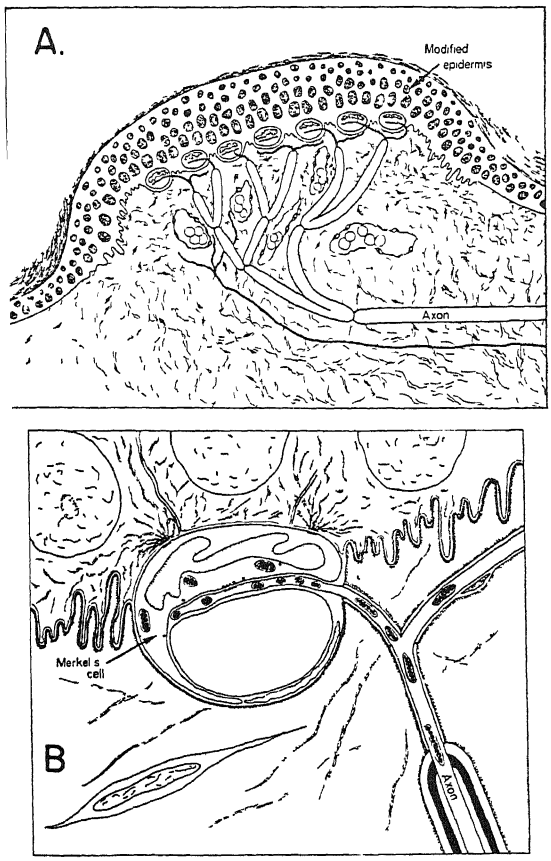


Fig 8 132 *A*, cross section of an epidermal dome with layer of nerve endings just inside the basement membrane Endings are derived from a single myelinated axon *B*, detail of the Merkel s cell at one of the nerve endings (Iggo 1966)

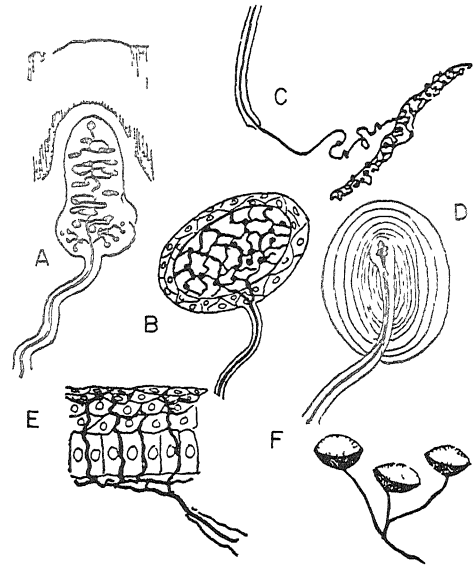


Fig 8 133 Various types of cutaneous receptors A Meissner s corpuscle (touch) B Krause s end bulb (cold) C Ruffini s end organs (warmth) D Pacinian corpuscle (deep pressure) E bar nerve endings in the cornea (pain) F Merkel s disks (touch) (Bainbridge and Menzie 1925)

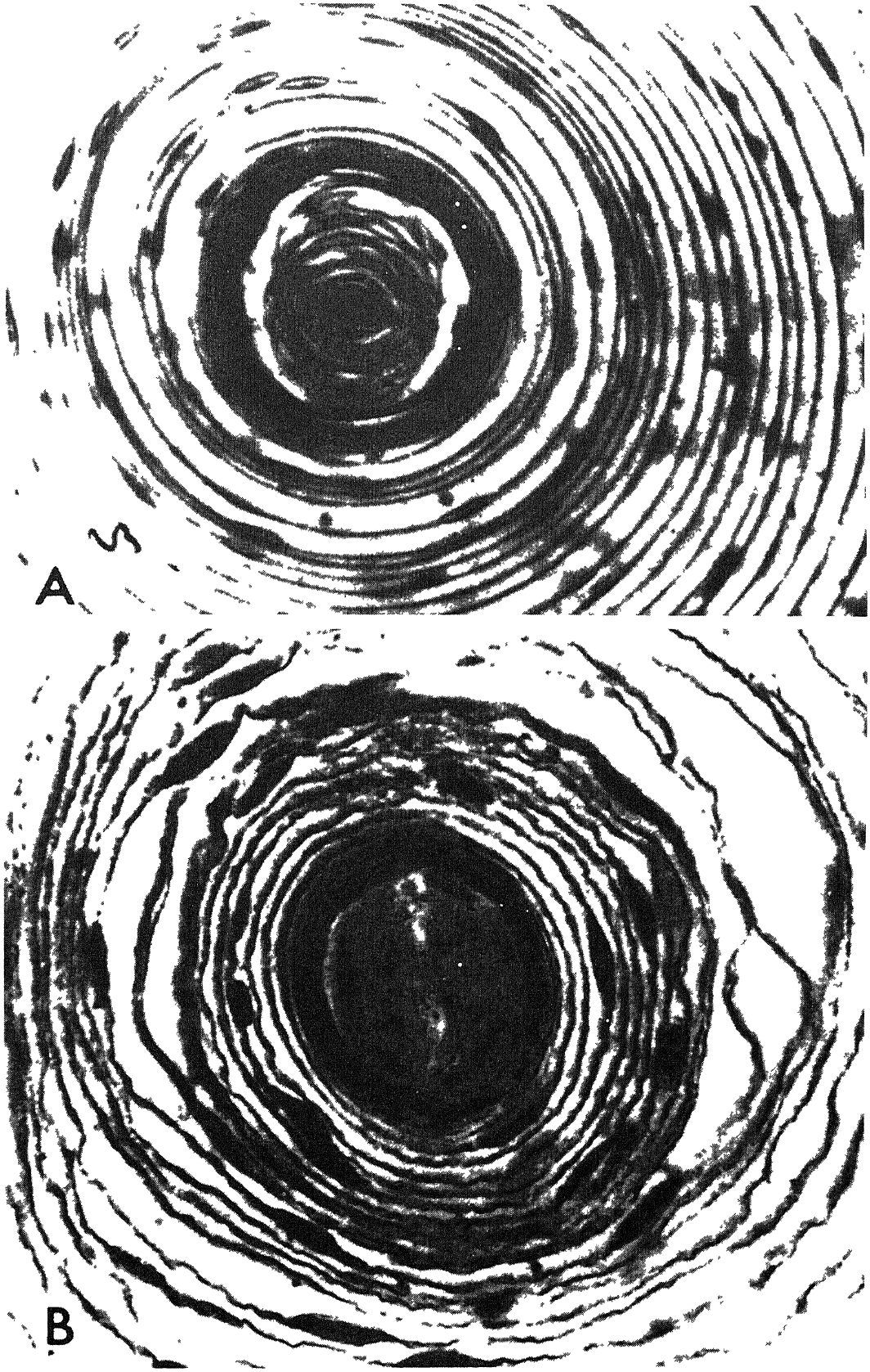
old, and are associated with receptive fields of as large as 5 cm ^{2} for a threshold level of stimulation These receptors are apparently served by myelinated fibers of up to 10 μ in diameter

Rapidly adapting units are found in nonhairy skin such as the footpad of the monkey These also show a discharge frequency which is relatively independent of the stimulus displacement Pacinian corpuscles (Fig 8 134) are associated with rapidly adapting responses to mechanical stimulation and may mediate this type of response when hair follicles are not involved They have also been associated with "deep touch"

The second type of receptor responds initially with a burst of activity and then adapts to a lower, sustained level of discharge This is sustained so long as the stimulation is continued for up to 5 min or even longer Recently, Iggo (1966) has associated the rapidly and slowly adapting receptor processes with specific anatomically defined afferent units

Fig 8 134 *A* unstained osmium-fixed transverse section of a cat Pacinian corpuscle as viewed by interference microscopy (× 750) The axial nerve terminal is oval in cross-section

B transverse section through a denervated Pacinian corpuscle of cat showing about 25 lamellae of the outer zone (× 900) Loss of circularity in the concentric outer lamellae and an apparent increase in the size of the lamellar nuclei are attributed to formol fixation and silver staining (Quilliam 1966)


A
B

He has associated the rapidly adapting receptors with the hair cells and divides these into three types The most sensitive are type D, characteristic of down hairs Their receptive fields are up to 2 cm ^{2} in area, and hence, many hairs must be capable of triggering a receptor response which converges on a single neural pathway It is probable that divergent connections also exist

Type G follicle units are those associated with "guard" hairs Ten or more of these hairs may be connected to a single neuron and in the 2-cm ^{2} area in which they are located other hairs will be connected to other neurons A kind of specialized guard hair termed a tylotrich follicle is found fairly commonly in the rabbit The type T-unit consists of an efferent neuron supplied by from 3 to 10 of these specialized guard hairs Each of the follicles is surrounded by a large venous sinus Electrophysiological studies of the monkey suggest that most of the rapidly adapting units found in this animal fall into the D and G classes The rate of discharge which occurs when a hair is bent apparently has little relation to the extent of bending When a threshold amount of bending has occurred, the discharge follows and if the hair is then maintained, perfectly undisturbed, in a fixed position, the unit adapts More rapid bending to a greater position of deflection serves only to reduce the latency of response The duration of response may be influenced by the number of hairs which are stimulated but rarely lasts much longer than 100 or 200 msec

The discharge of the slowly adapting units is very persistent In addition, the frequency of discharge varies with the amplitude of stimulation, the greater the displacement of the skin, the greater the discharge frequency Iggo has identified two types of slowly adapting units in hairy skin of the cat, rabbit, and monkey *Type I*, may produce a burst of impulses at rates in excess of 1000/sec when a smooth probe is drawn across a hairy skin surface This is a much higher discharge frequency than ever found with hair follicle units This type of response has been associated with the dome-shaped elevations of the skin which appear to contain Merkel cells as illustrated in Figure 8.132 The frequency of discharge increases with the amount of displacement of the tissue A mechanical stimulation of short duration, 1 or 2 msec, may cause a discharge of impulses which will last for as long as 10 msec For a fixed amount of displacement, the frequency of discharge is directly proportional to the displacement amplitude Similar results have been reported by Werner and Mountcastle (1968) for one class of receptors The type I units have a highly localized response and are relatively insensitive to lateral stretching of the skin They are normally quiescent at temperatures between 30° and 35° C in the absence of mechanical stimulation

Another differentiation in the nature of the response of tactual receptors relates to their ability to summate over time and area Pacinian corpuscles have been associated with responses which may be increased in magnitude by increasing either the duration or the area of stimulation There are other receptors which mediate a much more sharply defined response in both space and time (Verrillo, 1968) The latter are probably associated with the rapidly adapting receptors described above

The second class of slowly adapting receptors, *type II*, have a regular resting discharge in the absence of any mechanical stimulation When stimulation is applied, the initial response is of lesser amplitude than that of the type I units There is a reduction from the initial level following an adaptation period of several seconds, and with sustained stimulation a sustained level of discharge higher than the resting level may be measured These units are quite sensitive to lateral stretching of the skin So far, it has not been possible to associate these units, identified electrophysiologically, with anatomical receptor types Mountcastle and his colleagues (Werner and Mountcastle, 1968) have found some interesting differences in the relation between amplitudes of tactual stimulation and the response discharge frequency Some units have been observed (Fig 8 135) which show a linear relation between discharge frequency and tissue displacement Others show a linear relation between the logarithm of the response and the logarithm of the stimulus The slope of this linear relation is in the neighborhood of 0 5 and the response may, thus, be interpreted as a power function of stimulation with an exponent of 0 5 The linear relation referred to above is, of course, also a power function with an exponent 1 An interesting finding is that the results of scaling studies of tactual stimu-

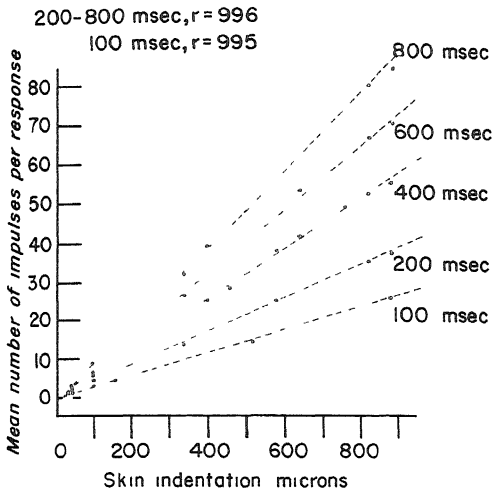


Fig 8 135 Number of impulses per response as a function of the depth of skin indentation for each of five observation times beginning at stimulus onset Recording from a single afferent fiber of the radial nerve of the Macaque Stimulus applied to the hand (Werner and Mountcastle 1968)

lation in human subjects reveal a functional relation between the magnitude of the stimulus and the scaled value, which is similar in form to the electrophysiological function derived from monkeys when the same region of the skin is stimulated

PERCEPTION OF TOUCH

Touch Localization. The location on the skin at which a tactual stimulus is applied can be identified fairly accurately by a human subject The accuracy varies, of course, with the region of the skin involved There is not a direct correspondence between the accuracy with which a point of stimulation may be localized and the minimum threshold for detection of stimulation The spatial relations among tactual receptors are obviously preserved in projections to the somaesthetic cortex, and there is some correspondence between localization accuracy and sensitivity The lips and the fingers, for which the tactual threshold is low, also permit highly accurate localization

The Two Point Threshold. A kind of discriminative localization threshold is provided by the test of the two-point threshold Two blunted compass points may be applied to the skin in various regions of the body The separation of the points may be varied in order to determine the minimum separation necessary for the subject to perceive the stimulus as two discrete points In such measurements, the subject is never aware as to whether the actual contact will be made by one point or two The experimenter is, thus, provided with an objective measure of the accuracy of the subject's report As one might expect, the minimum detectable spacing of two points is closely associated with the localization accuracy for various regions of the body's surface Two points may be discriminated as such when separated by only a little more than 2 mm at the fingertips, while a separation of as much as 6 or 7 cm may be necessary on the skin of the middle of the back, the upper arm, and the thigh

Localization accuracy may prove a useful basis for clinical diagnosis Certain cortical lesions are associated with a systematic displacement of the point associated with stimulation from the point actually stimulated on the same side of the body Certain forms of hysteria may be accompanied by a localization at a corresponding point on the opposite side of the body

The sense of touch is not restricted to receptors located in the skin So-called "deep touch" might more properly be called a response to pressure This is believed to be mediated by Pacinian corpuscles which are located in the dermus, tendons, periosteum, and other deep-seated structures

NEURAL PATHWAYS WHICH MEDIATE RESPONSES TO TOUCH

In the long ascending pathways of the spinal cord which conduct afferent signals from the skin and other parts of the body to the brain, there is a tendency for those fibers which mediate qualitatively similar responses to be grouped together Fibers from the skin and those from receptors located in deeper tissue may be located in the same portion of the ascending column when they carry information of a similar quality, such as touch The dorsal column-medial lemniscal system conveys information concerning the location and the nature of tactual stimulation, as well as its temporal variations It thus, appears to subserve precise discriminative functions The spinal thalamic system within the anterolateral tracts appears, on the other hand, to be more concerned with general aspects of sensation This tract has been associated with infor-

mation concerning the nature of the organism independent of the external world, although this kind of distinction is certainly not always valid Components of the spinal thalamic system project on the reticular formation in the medulla and the midbrain, as well as upon the intralaminar nuclei of the thalamus It is, thus, reasonable to associate it with arousal and vegetative processes (Mountcastle, 1961)

Projections on the somatic areas I and II of the cortex cannot be clearly differentiated, although there is a tendency for them to be associated with the lemniscal and the spinal thalamic components of the somatic afferent system, respectively The combined electrophysiological and behavioral studies of somaesthetic sensitivity in animals provide the basis of our current understanding of the system (Mountcastle et al , 1966)

Temperature Sense

Temperature is an index of relative heat Temperature, as such, is therefore probably not appropriately considered a physical stimulus, but it is a most satisfactory index of the possible effects of increases or decreases in the amount of heat in living tissue The temperature of various parts of the body may vary over a relatively narrow range without danger, but excessive heating or cooling can cause damage Temperatures above 45° C may result in protein denaturation (Hardy et al , 1952) This is associated with the sensation of pain Very low temperatures may also cause tissue damage but the anaesthetic effect of cooling under the appropriate conditions may conceal the fact that damage is being incurred and pain is not sensed until the cool tissue is reheated

Well within the range of noninjurious amounts of heat, receptors respond to relatively small amounts of temperature change The resulting sensations are of warmth or coolness These sensations are not painful and are qualitatively quite different from the sense of touch It is, therefore, reasonable to assume that they may be mediated by specialized receptor endings The receptor system for the detection of thermal variations provides a basis for the voluntary regulation of activity and the amount of clothing worn The system is also involved in reflex mechanisms which permit the close regulation of body temperature

RECEPTORS FOR THERMAL SENSITIVITY

Two types of end organs are characteristically associated with thermal sensation These are the Krause end organs (Fig 8 133*B*) associated with the discrimination of cooling, and the Ruffini end organs (Fig 8 133*C*) associated with the sensation of warming The Ruffini end organs may also be stimulated either electrically or mechanically and, if the stimulation is controlled appropriately, it will produce a sensation of warmth A Ruffini end organ is frequently found located near a deep plexus of blood vessels

METHODS OF STUDYING THERMAL SENSITIVITY

One of the earliest techniques for studying responsiveness to temperature change involved the use of thermal probes, often metallic containers with a blunt tip, the temperature of which could be controlled by the use of a hot or cold liquid contained in the device Investigation of the surface of the skin of human subjects has revealed regions with increased sensitivity to cold or warmth similar to those referred to above for touch Considerations of responsiveness to a variety of stimulus dimensions reveals that there are relatively fewer selectively responding spots for the detection of cold and tactile stimulation, as compared with warm and painful stimulation This may have evolved as a result of the fact that warm stimulation more frequently presages possible injury than does cold, and it is obvious that painful stimulation is typically an indication of potential injury

The thermal sensory system may be studied in terms of regulatory reflexes or the behavioral responses of animals Such types of investigations may be associated Reflex thermal regulation is treated in detail elsewhere

A third method of studying the temperature sense involves electrophysiological recording from single neural units in animal preparations or even in human subjects Some fairly recent data have been obtained with these techniques which help to verify the nature of the temperature-sensing system

Hensel (1966) and his colleagues have examined quantitatively the nature of the response of single fibers in a variety of organisms with changes in the temperature of stimulation of

an appropriate region of the surface of the skin The results make it clear that certain fibers are associated with cold receptors and others with warm receptors The discharge frequency in the cold response system is increased by any relatively abrupt decrease in temperature When there is no resting level of discharge, warming has no effect on these units A resting level of discharge may be inhibited by the presentation of a warm stimulus Stimulation of the region of the skin which produces a response to temperature change by a mechanical probe ordinarily produces no response or shows a very high threshold for mechanical stimulation The responses of a single cold fiber in the radial nerve of man are illustrated in Figure 8 136 It is evident that as the continuous trace which represents changing temperature drops, the discharge frequency of the fiber increases With subsequent increase in temperature the discharge frequency is again reduced Increased discharge frequency may be initiated for drops from a variety of initial temperatures, but the maximum rate is achieved for reductions in temperature from the region of 20° C

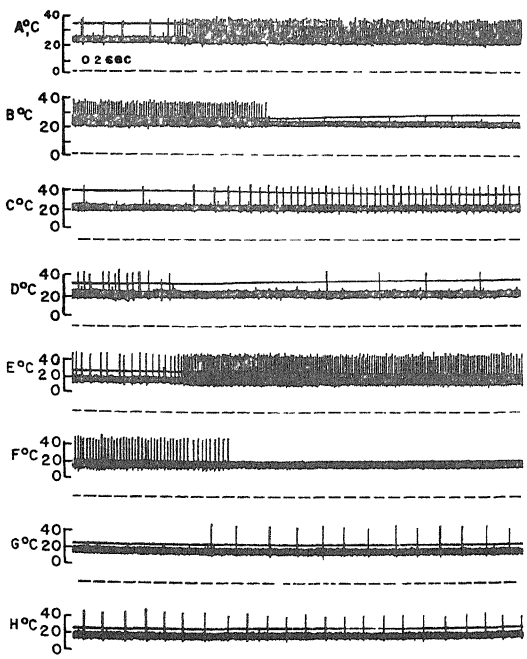


Fig 8 136 Afferent impulses in a single cold fiber in the radial nerve in man and temperature during cutaneous thermal stimulation (A) Start of cooling from 34° to 26° C (B) start of rewarming (C) start of cooling from 38° to 35° C (D) start of rewarming (E) start of cooling from 24° to 16° C (F) start of rewarming (G) 13 sec after rewarming has started (H) continued from record (G) (Hensel 1966)

Warm receptor systems are identical except that they respond in the opposite way to changes in temperature, rapid warming causes an increase in discharge frequency, while cooling either has no effect or inhibits a resting discharge frequency

Changes in response have been noted in a variety of animals (including the monkey) for a change in temperature as small as 0 2° from an initial value of 33° This compares reasonably well with the threshold change in temperature required for the elicitation of a warm or cold response in man

Both cold and warm receptor systems show a change in resting discharge frequency with a change in the sustained temperature level The resting discharge is a maximum for the warm system for temperatures between 38° and 43°, for the cold system between 15° and 34° The maximum discharge frequency ranges from 2 to 15 impulses/sec The characteristic response of a cold or a warm receptor system is seen only for relatively abrupt changes in temperature If the change is gradual, both the warm and the cold systems will show an increase in discharge frequency with increased temperature from below the range of maximum response up to that range, and a decrease in discharge frequency for increases in temperature above the optimum range

Stimulation of the Krause end organs by increase in temperature up to the vicinity of 45° C has been shown to be associated with the sensation of cold The electrophysiological correlate of this has been reported by Zotterman (1953) in cold fibers of the cat which are again activated with increase in temperature up to 45° C, and which continue to respond up to a maximum temperature of approximately 50° C The sensation of "paradoxical cold" can be observed in human subjects if the region of stimulation of the skin is selected carefully, and a stimulus temperature of approximately 45° or a little higher is employed

Certain sensory fibers have been found which show a change in discharge frequency for both mechanical stimulation and to reduction in temperature These are much less sen-

sitive to changes in temperature than the receptor systems we have been discussing, however Where the change in discharge frequency in impulses per second for every degree centigrade change in temperature may be of the order of 30 to 55 for units which respond only to thermal stimulation, this same index of response reduces to the order of 1 to 5 impulses/sec for each degree centrigrade change in temperature in those units which respond both to pressure and to cooling In human subjects, the stimulation of such receptors by mechanical means never produces a sensation of cold, but the application of a thermal stimulus may be perceived as a change in pressure It, therefore, seems reasonable to assume that these systems are in fact pressure-sensing systems

THEORETICAL CONSIDERATIONS

One hypothesis that has been suggested is the idea that the appropriate stimulus for a thermal receptor may be a gradient of temperature across its physical extent Thus, if the outer portion of the peripheral receptor is cooler or hotter than an inner portion, the result is a change in the discharge frequency of the system The possibility that this may be so has been refuted by showing that the introduction of a thermal stimulus, either above or below the receptor, has the same result in terms of change in discharge frequency It is, therefore, fairly clearly established that the appropriate stimulus is a change in temperature of the whole receptor termination

The question has been raised as to whether the specialized receptors in the skin for touch, temperature, and pain may all be basically chemoreceptors which respond to a transmitter substance which may be released by ancillary cells when an appropriate stimulus is applied (Mountcastle, 1966) The idea is an interesting one but it does not seem necessary for certain receptor types such as the Pacinian corpuscles (Loewenstein, 1966)

From our discussion of the warm and cold receptor systems studied by Hensel, it would seem that the coding of these dimensions of thermal response is accomplished by the specific responding neurons Some of the behavioral work of Kenshalo (1968) has been interpreted to suggest that coding may be in the form of increased activity for cold and decreased activity for warmth, but direct recording from single fibers does not accord with this opinion

Pain

Pain is an extremely important sensation by reason of the role it plays in the protection of the organism from hazardous agencies in the environment Reflex withdrawals mediated by a painful stimulus prevent or minimize burns, cuts, and all manner of potential injuries In addition to its direct relation to protective reflexes, pain is important for the learning of an organism to avoid future contacts with painful agents once encountered In our civilized society the survival value of pain is based on still an additional dimension of usefulness For the physician, the patient's report of pain, its location, its quality, its intensity, and its duration may play an extremely important role in medical diagnosis The clinical study of the significance of pain in injury and disease represents an extensive field which cannot be treated in detail here, however

STIMULI AND RECEPTORS

It has been extremely difficult to identify specific receptors which mediate pain There are very few tissues of the body which will not respond to damaging stimulation with pain Studies of single fibers render it clear that there are specific pathways which mediate pain sensations, but the receptors with which they are associated have not been clearly identified It is probable that a variety of cutaneous receptors and receptors for sensations other than pain within deeper tissue may provoke a painful response when the stimulus is extreme The best subjects for the study of pain are human subjects, by reason of their ability to provide a verbal confirmatory report of painful or uncomfortable sensations in response to stimulation Pain must be inferred in animals from observation of the behavioral correlates of the sensation in humans

Another problem arises from the fact that painful stimuli, except when very near threshold levels, usually are damaging to tissue Tissue which is damaged cannot be expected to continue to respond in the normal way Thus, in attempting to study pain in an animal preparation, the investigator may be altering the system in some fashion which will influence his results and his conclusions If he attempts to restrict stimulus intensity to a near threshold level, it may be virtually impos-

sible to determine that the effect of stimulation is indeed painful

The "pain" system is probably involved in nonpainful as well as painful sensations It has been associated with the sensation of taste of certain spicy foods (Lee, 1954), and as will be discussed below, may be associated with "tingling" or "itching" sensations which are not always unpleasant This additional role of the pain receptors and the neural pathways which mediate their response renders the study of the system in lower animals even more difficult

Pain Localization The localization of a painful stimulus is obviously of some importance if the sensation is to provide a basis for escape or withdrawal from the stimulus Localization of pain is reasonably good, but it does not have to be completely precise When the pain originates on the external body surface or in a limb, the withdrawal reflexes involve relatively gross movements of the whole body away from the source of stimulation, or flexion withdrawal of an entire limb There are certain circumstances where the localization of the pain as contrasted with the site of injury may be grossly in error For example, when there is injury to a nerve pathway which mediates pain sensation, the sensation may be projected to the origin of the pathway, such as a limb, although the actual injury may be in the region of the spinal cord The fact that a painful sensation may be projected to a region which has suffered no damage may in no way mitigate the extent of the pain, however

Another error in localization occurs with what is known as *referred pain* Damage or injury in a certain region may result in a painful sensation associated with a region somewhat removed from the locus of the injury This phenomenon differs from the projection of pain in that the region to which the pain is referred is not a more peripheral location along a nerve pathway passing through the locus of injury The pain associated with visceral injury or pathology is quite frequently referred to regions of the trunk somewhat removed from the focus of actual irritation The way in which erroneous localizations are associated with irritation in the gastrointestinal tract follows a fairly characteristic pattern, however, and the region where the pain is sensed may provide important information for diagnosis of the actual site of irritation

The Receptors for Pain. As indicated above, it has not been possible to make clear-cut identifications of pain receptors which uniquely serve the mediation of a pain response They probably are unspecialized endings in the skin and in other regions, and their existence is strongly supported by the existence of small diameter myelinated and unmyelinated fibers of slow conduction velocity which have been fairly clearly established to serve the sensation of pain (Clark et al , 1935) The regions of the body in which pain is experienced make it clear that these receptors must exist in skin, muscle, the viscera, the cornea, arteries but not veins, periosteum, and cancellous bone

Pain Stimulation Pain endings apparently respond indiscriminately to any stimulus, mechanical, chemical, or thermal when it is sufficiently intense The one generalization which can be made is that a painful stimulus, by reason of its nature or its intensity, is capable of causing injury An exception may be the stimuli which excite deep pain endings These may be located in muscles, tendons and joints, or in the periosteum Mechanical stimulation which would not cause pain in normal tissue may provoke a response following injury or infection Apparently, the deep pain endings greatly increase in sensitivity to mechanical and chemical stimulation as a sequel to injury or infection Muscle pain may result from prolonged muscular contraction Muscles become much more sensitive to pain during prolonged or even intermittent contraction when their supply of blood is reduced or interrupted This has been associated with the release of a chemical which is a product of muscle metabolism Its identity is unknown Pain does not occur with intermittent contraction of normal muscle, and it is relieved fairly rapidly in ischemic muscle when circulation is restored Even after the muscle contraction has ceased, the pain may continue in muscle which is deprived of blood until the normal supply is restored

For the scientific study of pain, it is desirable to employ quantitative methods of stimulus specification One procedure which has been widely used in stimulation of the skin involves controlled heating (Hardy et al , 1952) A region of the skin is blackened with India ink in order to equalize the absorptance of the skin of individuals with varying pigmentation A known amount of radiant heat from a calibrated lamp may then be applied in a restricted area for various durations Heating of the tissue may be measured fairly accurately

Pain is associated with the elevation of skin temperature to the neighborhood of 45° C, and a qualitative scale of increasing pain intensity may be related to increasing amounts of thermal stimulation Another method developed more recently for the study of chemical agents which may stimulate pain consists of the production of a blister base on the skin such that chemicals may be directly applied after the epidermis has been removed and any painful response associated with the creation of the blister has subsided (Keele, 1966) One problem with the use of some chemical agents is their indiscriminate excitation of a variety of receptors, in addition to those which may be uniquely associated with the pain response The blister base technique has proven useful for the comparative evaluation of a number of agents which may evoke a painful response in humans, nonetheless

PAIN AS A SENSATION

The Response to Pain The threshold for perception of pain based on quantitative studies with thermal stimulation has proven to be remarkably stable among individuals The perception threshold can be influenced by distracting agencies of a wide variety, e g , by hypnosis, by alcohol, and by certain drugs Ordinarily, however, the same amount of stimulation will result in a threshold perception in the same individual from one time to another, as well as in different individuals The nature of the reaction to painful stimulation, on the other hand, varies greatly from subject to subject It also varies greatly in a given subject depending on circumstances On a battlefield, serious injuries may be ignored and appear to cause the injured person remarkably little inconvenience Under other circumstances the same individual might be completely incapacitated by a lesser injury There is some question as to whether the effect is on the sensation of pain itself or on the nature of the reaction to that sensation (Melzack and Casey, 1968) Patients who have been given morphine, or others with frontal lobe lesions (Freeman and Watts, 1948) may be able to report pain sensation at the same stimulus level associated with the perceptual threshold in normal subjects, but the painful sensation no longer elicits the typical normal reactions It is as if the pain sensation alone "doesn't hurt!" Melzack and Casey have argued convincingly for a broader interpretation of "pain" than that possible when it is viewed as purely sensory It is certainly well established that the affective concomitants may be much more important in the determination of reaction to pain, and these affective concomitants are not always well correlated with the level of stimulation and, presumably, with the intensity of sensation

The Quality of Pain. Any one who has lived long enough to read this account is well aware of the great variety of painful sensations A pain may be characterized as "sharp" or "dull " It may be throbbing or sustained Sharp, bright, or pricking pain is usually of short duration A burning pain is more prolonged The "burning" quality of pain is not necessarily associated with a thermal stimulation Deep pain, originating in the tendons and joints or in the viscera may be associated with a drop in blood pressure, a slowing of the pulse, sweating, and nausea Such pain is often characterized as "sickening" and may even be accompanied by vomiting Cutaneous pain is more often associated with an increase in pulse rate and elevated blood pressure

Hyperalgesia The qualitative nature of a painful sensation may be altered following injury In an extended region around an injury to the skin, the threshold for sensation of pain may be slightly lowered and pain, once aroused, may be diffuse and extremely unpleasant (Lewis and Hess, 1933) The threshold may be changed relatively little or not at all, but stimuli which would normally create only mild irritation elicit a relatively intense and sustained pain response Similar effects in the skin may result from an injury to deeper tissue or underlying mucous membranes The extended region in which response to painful stimuli is changed is referred to as an area of *primary hyperalgesia* There is erythema or reddening of the skin as a result of vasodilatation The effect is not immediate upon injury, but may follow an interval during which there is little or no pain This suggests that the hyperalgesia does not result from continuing hyperactivity of damaged cells Rather, it is attributed to a release of some chemical substance in the injured region by the damaged tissue and in surrounding regions as a result of action of the nerve fibers which are involved

Another kind of hyperalgesia, *secondary hyperalgesia*, may be elicited by repeated electrical stimulation of a focal region of the skin

A change in response to painful stimulation occurs in a region around the point of stimulation this kind of hyperalgesia extends well beyond the region of erythema There is no reduction in threshold and, in fact, threshold may actually be increased slightly The sensation of pain, once aroused, is much more intense than normal The duration of secondary hyperalgesia is typically much shorter than that of primary hyperalgesia The latter may last for many days following severe injury

THE NEURAL PATHWAYS FOR PAIN

The early work of Clark et al (1935), showed that the high threshold afferent units associated with pain sensation are to be found among small myelinated and nonmyelinated fibers, slowly conducting C-fibers The work of Brown and Iggo (1967) has established that the majority of small myelinated fibers are excited by stimuli which can hardly be considered noxious The problem remains of determining which fibers are specifically associated with the conduction of pain information Some method must be conceived of selectively stimulating pain endings in the skin of experimental animals Perhaps the only solution, as suggested by Iggo, is a systematic effort to catalog the stimulus situations which give rise to pain in man and then identify those fibers and hopefully, receptors, which only respond to such stimuli when they are applied to animals It is clear that all small, nonmyelinated fibers are not exclusively pain pathways in the skin, muscle, or viscera The problem of selective study of pain pathways arises from the fact that a diversity of pathways in addition to those which are exclusively assigned to pain are brought into play by any stimulus which will elicit a pain response Even certain pathways which respond only to relatively high levels of stimulation may not mediate pain (Iggo, 1957, Titchen, 1958)

Combined Sensations

Tickling and itching have been associated with the stimulation of both touch and pain receptors A tickling sensation may not be detectable after injury to the lateral spinal thalamic tract, although stimuli can elicit sensations of touch On the other hand, when the sensation of touch is reduced or lost during moderate ischemia there is also no tickling sensation In this case, the sensitivity to pain remains If the lateral spinal thalamic tract is primarily concerned with pain response, then it would appear that tickling requires function of both the touch and pain receptor systems

Itching may be determined exclusively by the pain system (Rothman, 1943) or may, in some circumstances, depend upon both the touch and pain systems Spontaneous itching may depend upon the pain system exclusively It is unalleviated by relative ischemia which reduces or eliminates touch sensation Itching is believed to result from stimulation by a chemical mediator which is released when cells are damaged It may also be elicited by mechanical or electrical stimulation (Arthur and Shelley, 1959) Any agent which eliminates pain will also eliminate itching

The difficulty in identifying specific receptors for various cutaneous sensations and the evidence that nonspecified endings may subserve a variety of sensations have sometimes raised the question of whether the various skin sensations are properly considered independently The differential response of these various sensations would suggest that they do represent independent systems For example, touch is eliminated first by ischemia or reduced blood flow On the other hand, with the administration of cocaine, the sensation of cold, warmth, pain, and touch are lost in that order Local cooling of the skin results in loss initially of the sensation of cold, followed by touch, pain, and finally warmth

BIBLIOGRAPHY

ABRAHAMS H D KRAKAUER AND K M DALLENBACH Gustatory adaptation to salt *Amer J Psychol* 49 462-469, 1937

ADES H W AND H ENGSTROM Form and innervation of the vestibular epithelia In *The Role of the Vestibular Organs in the Exploration of Space* Nat Aeronautics Space Admin NASA SP-77 23-41 1965

ADES, H W, H ENGSTRÖM, AND J E HAWKINS Structure of inner ear sensory epithelial cells in relation to their functions BUMED Project MR 005 13 2005, Report No 1 Pensacola Fla Naval School of Aviation Medicine, 1962

ADLER F H *Physiology of the Eye Clinical Application* Ed 3 St Louis Mosby 1959

ADLER, F H *Physiology of the Eye* Ed 4 St Louis Mosby 1965

ADRIAN E D Discharges from vestibular receptors in the cat *J Physiol (London)* 101 389-407, 1943

ADRIAN E D Olfactory discrimination *Annee Psychol* 50 107-113 1951

ADRIAN, E D *Olfaction and Taste Proc 1st Int Symp*, Wenner Gren Center edited by Y Zotterman New York MacMillan 1963

ADRIAN E D AND C LUDWIG Nervous discharges from the olfactory organs of fish *J Physiol (London)* 94 441-460 1938

ALLANSON J T AND I C WHITFIELD The cochlear nucleus and its relation to theories of hearing In *The Third London Symposium on Information Theory* edited by C Cherry London Butterworth 1955 p 269-286

AMOORE J F Psychophysics of odor *Cold Spring Harb Symp* receptors *Perfum Essent Oil Rec Yb* 43 321-323 330 1952

AMOORE J E Stereochemical theory of olfaction *Nature (London)* 198 271-272 1963

AMMORE J E Psychophysics of odor *Cold Spring Harb Symp Quant Biol* 30 623-637 1965

AMOORE J E J W JOHNSTON AND M RUBIN The stereochemical theory of odor *Scient Amer* 210 (Suppl 2) 42-49 1964

AMOORE J E D VENSTROM AND A R DAVIS Measurement of specific anosmia *Percept Mot Skills Res Exch* 26 143-164 1968

ANDERSEN H T, M FUNAKOSHI AND Y ZOTTERMAN Electrophysiological responses to sugars and their depression by salt In *Olfaction and Taste*, edited by Y Zotterman New York Macmillan 1963

ANDERSON R J Taste thresholds in stimulus mixtures *Microfilm Abstr* 10 (Suppl 4) 287-288 1950

APTER J T Eye movements following strychninization of the superior colliculus of cats *J Neurophysiol* 9 73-86 1946

ARDEN, G B, AND U SODERBERG The transfer of optic information through the lateral geniculate body of the rabbit In *Sensory Communication* edited by W A Rosenblith New York Wiley 1961 p 521-544

ARDEN G B AND R A WEALE Nervous mechanisms and dark adaptation *J Physiol (London)* 125 417-426 1954

ARMINGTON J C E P JOHNSON AND L A RIGGS The scotopic A wave in the electrical response of the human retina *J Physiol (London)* 118 289-298, 1952

ARTHUR R P AND W B SHELLEY The peripheral mechanism of itch in man In *Pain and Itch Nervous Mechanisms* edited by G E W Wolstenholme and Maeve O Connor Boston Little Brown 1959 p 84-97

ARY L B, M J TREMAINE AND F L MONZINGO The numerical and topographical relations of taste buds to human circumvallate papillae throughout the life span *Anat Rec* 64 9-25 1935

ASCHAN, G Different types of alcohol nystagmus *Acta Otolaryng* (Suppl 140) 69-78 1958

BAINBRIDGE, F A AND J A MENZIE *Essentials of Physiology* Ed 5 edited and revised by C L Evans New York Longmans, 1925

BAKER H D The course of foveal light adaptation measured by the threshold intensity increment *J Opt Soc Amer* 39 172-179 1949

BAKER H D The instantaneous threshold and early dark adaptation *J Opt Soc Amer* 43 798-803, 1953

BAKER, H D Initial stages of dark and light adaptation *J Opt Soc Amer* 53 98-103 1963

BARLOW, H B Dark adaptation a new hypothesis *Vision Res* 4 47-58, 1964

BARLOW, H B Trigger features adaptation and economy of impulses In *Information Processing in the Nervous System*, edited by K N Leibovic Gottingen Heidelberg Springer 1969, chap 11, p 209-226

BARLOW, H B, C BLAKEMORE AND J D PETTIGREW The neural mechanism of binocular depth discrimination *J Physiol (London)* 193 327-342 1967

BARLOW, H B R FITZHUGH AND S W KUFFLER Change of organization in the receptive fields of the cat's retina during dark adaptation *J Physiol (London)* 137 338-354, 1957

BARTLETT N R Thresholds as dependent on some energy relations and characteristics of the subject In *Vision and Visual Perception* edited by C H Graham New York Wiley, 1965a, chap 7, p 154-184

BARTLETT N R Dark Adaptation and light adaptation In *Visual Perception* ed by C H Graham New York Wiley 1965b chap 8 p 185-207

BAUMGARTNER, G, J L BROWN AND A SCHULZ Visual motion detection in the cat *Science* 146 1070-1071, 1964

BAUMGARTNER, G, J L BROWN AND A SCHULZ Responses of single units of the cat visual system to rectangular stimulus patterns *J Neurophysiol* 28 1-18 1965

BECK L H AND W R MILES Some theoretical and experimental relationships between infrared absorption and olfaction *Science* 106 511 1947

BEIDLER L M Dynamics of taste cells In *Olfaction and Taste* edited by Y Zotterman New York Macmillan 1963 p 133-145

BEIDLER L M M S NEJAD R L SMALLMAN AND H TATEDA Rat taste cell proliferation *Fed Proc* 19 302 1960

BENSON A J AND M A BODIN Interaction of linear and angular accelerations on vestibular receptors in man *Aerospace Med* 37 144-154 1966

BERSON E L AND E B GOLDSTEIN Recovery of the human early receptor potential during dark adaptation in hereditary retinal disease *Vision Res* 10 219-226 1970

BISHOP P O AND R DAVIS Bilateral interaction in the lateral geniculate body *Science* 118 241-243 1953

BIZZI E Discharge of frontal eye field neurons during saccadic and following eye movements in unanesthetized monkeys *Exp Brain Res* 6 69-80 1968

BIZZI E AND P H SCHILLER Single unit activity in the frontal eye fields of unanesthetized monkeys during eye and head movement *Exp Brain Res* 10 151-158 1970

BLACKWELL H R AND O M BLACKWELL Blue monocone monochromacy a new color vision defect *J Opt Soc Amer* 47 338 1957

BLACKWELL H R, AND O M BLACKWELL Rod and cone receptor mechanisms in typical and atypical congenital achromatopsin *Vision Res* 1 62-107 1961

BLAKEMORE C AND F W CAMPBELL On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images *J Physiol (London)* 203 237-260 1969

BLAKEMORE C AND P SUTTON Size adaptation A new aftereffect *Science* 166 245-247 1969

BLOMQUIST A J, R M BENJAMIN AND R EMMERS Thalamic localization of afferents from the tongue in squirrel monkey (*Saimiri scureus*) *J Comp Neurol* 118 77-88 1962

BORING E G *Sensation and Perception in the History of Experimental Psychology* New York Appleton 1942

BORNSTEIN W S Cortical representation of taste in man and monkey *Yale J Biol Med* 13 133-136, 1940

BOYNTON, R M, AND D N WHITTEN Visual adaptation in monkey cones Recordings of late receptor potentials *Science* 170 1423-1426 1970

BRINDLEY G S *Physiology of the Retina and the Visual Pathway* London Arnold, 1960

BROADBENT D E AND M GREGORY Accuracy of recognition for speech presented to the right and left ears *Quart J Exp Psychol* 16 359-360 1964

BRODAL, A The cranial nerves anatomy and anatomicoclinical correlations Springfield Ill Thomas, 1959

BROWN, A G, AND A IGGO A quantitative study of cutaneous receptors and afferent fibres in the cat and rabbit *J Physiol (London)* 193 707-733 1967

BROWN, J L Orientation to the vertical during water immersion *Aerospace Med* 32 209-217 1960

BROWN, J L The structure of the visual system In *Vision and Visual Perception*, edited by C H Graham New York Wiley 1965a

BROWN, J L Flicker and intermittent stimulation In *Vision and Visual Perception*, edited by C H Graham New York Wiley, 1965b

BROWN, J L, C H GRAHAM, H LEIBOWITZ, AND H B RANKEN Luminance thresholds for the resolution of visual detail during dark adaptation *J Opt Soc Amer* 43 197-202, 1953

BROWN, J L, J W METZ AND J R YOHMAN A test of scotopic suppression of the photopic process *J Opt Soc Amer* 59 1677-1678, 1969

BROWN J L, AND G MUELLER Brightness discrimination and brightness contrast In *Vision and Visual Perception*, edited by C H Graham New York Wiley 1965

BROWN, K S, AND R R ROBINETTE No simple pattern of inheritance in ability to smell solutions of cyanide *Nature (London)* 215 406-408, 1967

BROWN, K T The electroretinogram Its components and their origins *Vision Res* 8 633-677 1968

BROWN K T AND T N WIESEL Intraretinal recording in the unopened cat eye *Amer J Ophthal* 46 91–98 1958

BROWN K T AND T N WIESEL Intraretinal recording with micro pipette electrodes in the intact cat eye *J Physiol (London)* 149 537–562 1959

BROWN P K AND G WALD Visual pigments in human and monkey retinas *Nature (London)* 200 37–43 1963

BROWN P K, AND G WALD Visual pigments in single rods and cones of the human retina *Science* 144 145–151 1964

BRUST CARMONA H H KASPRZAK AND E L GASTEIGER Role of the olfactory bulbs in X ray detection *Radiat Res* 29 354–361 1966

BUJAS Z Quelques remarques sur le contraste et l inhibition a la suite d excitations gustatives simultanees *C R Soc Biol (Paris)* 116 1304–1306 1934

CAJAL S R *Textura del Sistema Nervioso del Hombre y de los Vertebrados* Madrid Moya 1904

CAMPBELL F W AND J J KULIKOWSKI Orientational selectivity of the human visual system *J Physiol (London)* 187 437–445 1966

CAMPBELL F W J NACHMIAS AND J JUKES Spatial frequency discrimination in human vision *J Opt Soc Amer* 60 555–559 1970

CAMPBELL J F D BINDRA, H KREBS AND R P FERENCHAK Responses of single units of the hypothalamic ventromedial nucleus to environmental stimuli *Physiol Behav* 4 183–188 1969

CAUNA N Fine structure of the receptor organs and its probable functional significance In *Touch, Heat and Pain*, edited by A V S deReneck and J Knight Boston Little, Brown 1966, p 117–127

CAUNA N Light and electron microscopol structure of sensory end organs in human skin In *The Skin Senses*, edited by D R Kenshale Springfield, Ill Thomas, 1968, chap 2 p 15–29

CHAPANIS A AND R A McCLEARY Interposition as a cue for the perception of relative distance *J Gen Psychol* 48 113–132 1953

CHOW, K L Numerical estimates of the auditory central nervous system of the rhesus monkey *J Compt Neurol* 95 159–175, 1951

CHOW, L K AND P J HUTT The 'association cortex" of Macaca mulatta a review of recent contributions to its anatomy and functions *Brain* 76 625–677 1953

CLARK B Thresholds for the perception of angular acceleration in man *Aerospace Med* 38 443–450, 1967

CLARK, B AND J D STEWART Comparison of sensitivity for the perception of bodily rotation and the oculogyral illusion *Percept Psychophys* 3 253–256, 1968

CLARK D, J HUGHES AND H S GASSER Afferent function in the group of nerve fibers of slowest conduction velocity *Amer J Physiol* 114 69–76, 1935

CLARK W E L The laminar pattern of the lateral geniculate nucleus considered in relation to colour vision *Documenta Ophth* III 57–64 1949

CLARK, W E L AND G G PENMAN The projection of the retina in the lateral geniculate body *Proc Roy Soc (London)* [*Biol*] *Ser B* 114 291–313 1934

COGAN D G, AND E V KINSEY The cornea V Physiological aspects *Arch Ophthal (Chicago)* 28 661–669, 1942

COHEN, A I The fine structure of the extrafoveal receptors of the Rhesus monkey *Exp Eye Res* 1 128–136, 1961

COHEN M J, S HAGIWARA AND Y ZOTTERMAN The response spectrum of taste fibres in the cat A single fibre analysis *Acta Physiol Scand* 33 316–332, 1955

COLEMAN P D, AND A H RIESEN Environmental effects on cortical dendritic fields I Rearing in the dark *J Anat* 102 363–374, 1968

COLLINS, C C, AND F SAUNDERS Pictorial display by direct electrical stimulation of the skin *J Biomed Syst* 1 3–16, 1970

COLLINS, W E Vestibular responses from figure skaters *Aerospace Med* 37 1098–1104, 1966

CORREA, M J W C HIXSON, AND J I NIVEN On predictive equations for subjective judgments of vertical and horizon in a force field *Acta Otolaryng (Stockholm)* (Suppl 230) 1–20, 1968

CRAGG, B G Centrifugal fibres to the retina and olfactory bulb and composition of the supraoptic commissures in the rabbit *Exp Neurol* 5 406–427, 1962

CRAMPTON, G H Response of single cells in cat brain stem to angular acceleration in the horizontal plane In *The Role of the Vestibular Organs in Space Exploration* Nat Aeronautics Space Admin NASA SP 77 85–96 1965

CRAMPTON G H AND F A YOUNG The differential effect of a rotary visual field on susceptible and nonsusceptible to motor sickness *J Compl Physiol Psychol* 46 451–453 1953

CRAWFORD B H Visual adaptation in relation to brief conditioning stimuli *Proc Roy Soc (London)* [*Biol*] *Ser B* 134 283–302 1947

CROSBY E C T HUMPHREY AND E W LAVER *Correlative Anatomy of the Nervous System* New York Macmillan 1962

DALLOS, P Z G SCHOENY AND M A CHEATHAM Cochlear summating potentials composition *Science* 170 641–644 1970

DARTNALL H J A *The Visual Pigments* London Methuen 1957

DAVIES J T A theory of the quality of odors *J Theor Biol* 8 1–7 1965

DAVIS H Psychophysiology of hearing and deafness In *Handbook of Experimental Psychology* edited by S S Stevens New York Wiley 1951 p 1116–1142

DAVIS, H *Olfaction and Taste* edited by Y Zotterman New York Macmillan 1963 p 145–148

DAVIS H Slow cortical response evoked by acoustic stimuli *Acta Otolaryng (Stockholm)* 59 179–185 1965a

DAVIS H Guide for the classification and evaluation of hearing handicaps in relation to the international audiometric zero *Trans Amer Acad Ophthal Otolaryng* 69 740–751 1965b

DAVIS H R W BENSON W P COVELL C FERNANDEZ R GOLDSTEIN Y KATSUKI J P LEGOUIX D R McAULIFFE AND I KATSUKI Acoustic trauma in the guinea pig *J Acoust Soc Amer* 25 1180–1189 1953

DAVIS H C T MORGAN, J E HAWKINS, JR R GALAMBOS AND F W SMITH Temporary deafness following exposure to loud tones and noise *Acta Otolaryng (Stockholm)* (Suppl 88) 57 pp 1950

DAVSON H *The Physiology of the Eye* Philadelphia Blakiston 1949

DAVSON, H The intra ocular pressure and the mechanism of formation of the intra-ocular fluid *Ophthal Lit (Denver)* 4 3–13 1950

DAVSON, H *The Eye* New York Academic Press 1962

DE LANGE H Research into the dynamic nature of the human fovea cortex systems with intermittent and modulated light I Attenuation characteristics with white and colored light *J Opt Soc Amer* 48 777–784 1958

DE LORENZO, A J Studies on the ultrastructure and histophysiology of cell membranes nerve fibers and synaptic junctions in chemoreceptors In *Olfaction and Taste*, edited by Y Zotterman New York Macmillan 1963, p 5–17

DEPALMA J J AND E M LOWRY Sine wave response of the visual system II Sine wave and square wave contrast sensitivity *J Opt Soc Amer* 52 328–335, 1962

DE ROBERTIS E J Morphogenesis of retinal rods, an electron microscope study *J Biophys Biochem Cytol* 2 209–218 1956

DESMEDT, J E Auditory evoked potentials from cochlea to cortex as influenced by activation of the efferent olivo cochlear bundle *J Acoust Soc Amer* 34 1478–1496 1962

DE VALOIS, R L Color vision mechanisms in the monkey *J Gen Physiol* 43 (Suppl 6) 115–128, 1960

DEVALOIS R Behavioral and electrophysiological studies of primate vision In *Contributions to Sensory Physiology*, edited by W D Neff New York Academic Press, 1965, vol I p 137–178

DEVALOIS, R L, G H JACOBS, AND A E JONES Effects of increments and decrements of light on neural discharge rate *Science* 136 986–987, 1962

DEVALOIS, R L, C J SMITH, A J KAROLY, AND S T KITAI Electrical responses of primate visual system I Different layers of macaque lateral geniculate nucleus *J Comp Physiol Psychol* 51 662–668, 1958

DEVRIES, H L AND P SCHIERBECK The minimum perceptible angular velocity under various conditions *Pract Otorhinolaryng (Basel)* 15 65–72, 1953

DEVRIES, H, AND M STUIVER The absolute sensitivity of the human sense of smell In *Sensory Communication*, edited by W A Rosenblith Cambridge M I T Press 1961, p 159–167

DINC H I, AND J C SMITH Role of the olfactory bulbs in the detection of ionizing radiation by the retina *Physiol Behav* 1 139–144, 1966

DOHLMAN, G F Histochemical studies of vestibular mechanisms

In *Neural Mechanisms of the Auditory and Vestibular Systems* edited by G L Rasmussen and W F Windle Springfield Ill Thomas 1960

DOWLING J E Foveal receptors of the monkey retina fine structure *Science* 147 57–59 1965

DOWLING J E The site of visual adaptation *Science* 155 273 1967

DOWLING J E Organization of vertebrate retinas *Invest Ophthal* 9 655–680 1970

DOWLING, J E AND R HUBBARD Effects of brilliant flashes on light and dark adaptation *Nature (London)* 199 972–975 1963

DUENSING F AND K P SCHAEFER Die Aktivitat einselner Neurone im Bereich der Vestibulariskerne bei Horizontalblesch leunigungen unter besonderer Berucksichtigung des vestibularen Nystagmus *Arch Psychiat u Z Neurol* 198 225–252 1958

DUKE ELDER S *Textbook of Ophthalmology* Ed 2 St Louis Mosby 1937 vol I

DUKE ELDER, S *System of Ophthalmology* St Louis Mosby 1961

DYSON G M The Raman effect and the concept of odour *Perfum Essent Oil Rec Yb* 28 13–19 1937

ECKEL W Elektrophysiologische und histologische Untersuchungen im Vestibulariskerngebiet bei Drehreizen *Arch Ohr- Nas u KehlkHeilk* 164 487–513, 1954

EFRON R The minimum duration of a perception *Neuropsychologia* 8 57–63 1970

ELSBERG C A AND J I LEVY The sense of smell I A new and simple method of quantitative olfactometry *Bull Neurol Inst N Y* 4 5–9 1935

ENGSTRÖM, H The innervation of the vestibular sensory cells *Acta Otolaryng (Stockholm)* (Suppl 163) 30–40 1961

ENGSTRÖM, H The first order vestibular neuron In *Fourth Symposium on the Role of Vestibular Organs in Space Exploration* Nat Aeronautics Space Admin NASA SP 187 123–134, 1968

ENGSTRÖM H H W ADES AND J E HAWKINS JR Structure and functions of the sensory hairs of the inner ear *J Acoust Soc Amer* 34 1356–1363 1962

ENGSTRÖM, H H W ADES, AND J E HAWKINS, JR Cellular pattern nerve structures and fluid spaces of the organ of Corti In *Contributions to Sensory Physiology*, edited by W D Neff New York Academic Press 1965a vol I, p 1–37

ENGSTRÖM H H W ADES AND J E HAWKINS The vestibular sensory cells and their innervation (In *Modern Trends in Neuromorphology* edited by J Szentagothai) *Symp Biol Hung* 5 21–41, 1965b

ENGSTRÖM, H, AND J WERSALL The ultrastructural organization of the organ of Corti and of the vestibular sensory epithelia *Exp Cell Res* (Suppl 5) 460 1958

ERICKSON, R P Duplexity theory of taste *Science* 147 890, 1965

ERICKSON R P *The Chemical Senses and Nutrition*, edited by M R Kare and O Maller Baltimore Johns Hopkins Press 1967 p 313

ERULKAR, S D, G L GERSTEIN AND R A BUTLER. Transmembrane potentials from cat cochlear nucleus in response to pure and frequency modulated (FM) tonal stimuli *Fed Proc* 25 463, 1966

ERULKAR, S D, P G NELSON, AND J S BRYAN Experimental and theoretical approaches to neural processing in the central auditory pathway In *Contributions to Sensory Physiology*, edited by W D Neff New York Academic Press, 1968 vol 3 p 150–189

ERULKAR, S D J E ROSE, AND P W DAVIES Single unit activity in the auditory cortex of the cat *Johns Hopkins Med J* 99 55–86 1956

EWALD, J R *Physiologische Untersuchungen uber das Endorgan des Nervus octavus* Wiesbaden, Germany Bergmann, 1892

FERNANDEZ C The innervation of the cochlea (guinea pig) *Laryngoscope* 61 1152–1172, 1951

FEX, J Auditory activity in centrifugal and centripetal cochlear fibres in cat a study of a feedback system *Acta Physiol Scand*, 55 (Suppl 189) 1–68 1962

FICK, A *Lehrbuch der Anatomie und Physiologie der Sinnesorgane* Schauenberg, Germany Lahr, 1864, vol 1, p 85

FISHMAN, I Y Single fiber gustatory impulses in rat and hamster *J Cell Physiol* 49 319–334 1957

FLOCK A Transducing mechanisms in the lateral line canal organ receptors *Cold Spring Harb Symp Quant Biol* 30 133–144, 1965

FOSTER, D E H SCOFIELD AND K M DALLENBACH An olfactorium *Amer J Psychol* 63 431–440 1950

FRANCOIS J *Heredity in Ophthalmology* St Louis Mosby 1961

FRANK M AND C PFAFFMANN Taste nerve fibers a random distribution of sensitivities to four tastes *Science* 164 1183–1185 1969

FREEMAN W AND J W WATTS Pain mechanisms and the frontal lobes a study of prefrontal lobotomy for intractable pain *Ann Intern Med* 28 747 1948

FRENCH N R AND J C STEINBERG Factors governing the intelligibility of speech sounds *J Acoust Soc Amer* 19 90–119 1947

FRISHKOPF L S AND M H GOLDSTEIN JR Responses to acoustic stimuli from single units in the eighth nerve of the bullfrog *J Acoust Soc Amer* 35 1219–1228 1963

GACEK R R Efferent component of the vestibular nerve In *Neural Mechanisms of the Auditory and Vestibular Systems*, edited by G L Rasmussen and W F Windle Springfield Ill Thomas 1961

GACEK R R Anatomical evidence for an efferent vestibular pathway In *Third Symposium on the Role of the Vestibular Organs in Space Exploration* Nat Aeronautics Space Admin NASA SP 152 203–211 1967

GALAMBOS R Suppression of auditory nerve activity by stimulation of efferent fibers to cochlea *J Neurophysiol* 19 424–437 1956

GALAMBOS R Studies of the auditory system with implanted electrodes In *Neural Mechanisms of the Auditory and Vestibular Systems*, edited by G L Rasmussen and W F Windle Springfield Ill Thomas, 1960

GALAMBOS R, AND H DAVIS The response of single auditory nerve fibers to acoustic stimulation *J Neurophysiol* 6 39–57, 1943

GALAMBOS R AND H DAVIS Inhibition of activity in single auditory nerve fibers by acoustic stimulation *J Neurophysiol* 7 287–303 1944

GANZ L Sensory deprivation and visual discrimination In *Handbook of Sensory Physiology* edited by H Lukas Teuber In press 1971 vol 8

GARRETT M, AND J FODOR Psychological theories and linguistic constructs In *Verbal Behavior and General Behavior Theory*, edited by T R Dixon and D L Horton New Jersey Prentice Hall 1968

GASSER H S Effect of method of leading on the recording of the nerve fiber spectrum *J Gen Physiol* 43 927 1960

GAZZANIGA M S The split brain in man *Sci Amer* 217 24–29 1967

GELDARD, F A *The Human Senses* New York Wiley 1953

GERNANDT B Response of mammalian vestibular neurons to horizontal rotation and caloric stimulation *J Neurophysiol* 12 173–184 1949

GESTELAND, R C J Y LETTVIN W H PITTS, AND A ROJAS Odor specificities of the frog's olfactory receptors In *Olfaction and Taste*, edited by Y Zotterman New York Macmillan, 1963 p 19–34

GIBSON J J *The Senses Considered as Perceptual Systems* Boston Houghton Mifflin 1966 p 355

GILINSKY A S Orientation specific effects of patterns of adapting light on visual acuity *J Opt Soc Amer* 58 13–18 1968

GILINSKY A S AND R S DOHERTY Interocular transfer of orientational effects *Science* 164 454–455, 1969

GLEES, P AND W E L CLARK The termination of optic fibers in the lateral geniculate body of the monkey *J Anat* 75 295–308 1941

GOGEL, W C, B O HARTMAN, AND G S HARKER The retinal size of a familiar object as a determiner of apparent distance *Psychol Monogr* 70 1–16 1957

GOLDSTEIN, E B, AND E L BERSON Cone dominance of the human early receptor potential *Nature (London)* 222 1272–1273, 1969

GOLDSTEIN E B AND E L BERSON Rod and cone contributions to the human early receptor potential *Vision Res* 10 207–218, 1970

GOURAS P Identification of cone mechanisms in monkey ganglion cells *J Physiol (London)* 199 533–547, 1968

GOURAS, P Electroretinography Some basic principles *Invest Ophthal* 9 557–569, 1970

GRAHAM, A B (ed) *Sensorineural Hearing Processes and Disorders* Boston Little, Brown 1966

GRAHAM, C H Some fundamental data In *Vision and Visual Per*

ception edited by C H Graham New York Wiley 1965a chap 4 p 68-80
GRAHAM C H Visual space perception In *Vision and Visual Perception*, edited by C H Graham New York Wiley 1965b chap 18 p 504-547
GRAHAM, C H, K E BAKER M HECHT AND V V LLOYD Factors influencing thresholds for monocular movement parallax *J Exp Psychol* 38 205-223 1948
GRAHAM, C H, AND J L BROWN Color contrast and color appearances Brightness constancy and color constancy In Vision and Visual Perception edited by C H Graham New York Wiley 1965 chap 16 pp 452-478
GRAHAM C H H G SPERLING Y HSIA AND A H COULSON The determination of some visual functions of a unilaterally color blind subject methods and results *J Psychol* 51 3-32 1961
GRANIT R Centrifugal and antidromic effects on ganglion cells of retina *J Neurophysiol* 18 388-411, 1955
GRAY, R F Functional relationships between semicircular canal function and otolith function *Aerospace Med* 31 413 1960
GRAYBIEL A, B CLARK AND J J ZARRIELLO Observations on human subjects living in a 'slow rotation room for periods of two days *Arch Neurol (Chicago)* 3 55-73 1960
GRAYBIEL, A AND E F MILLER The otolith organs as a primary etiological factor in motion sickness with a note on "off vertical rotation In *Fourth Symposium on the Role of the Vestibular Organs in Space Exploration* Nat Aeronautics Space Admin, NASA SP 187 53-64, 1968
GREENWOOD D D, AND N MARUYAMA Excitatory and inhibitory response areas of auditory neurons in the cochlear nucleus *J Neurophysiol* 28 863-892, 1965
GROSS C G, D B BENDER AND C E ROCHA MIRANDA Visual receptive fields of neurons in inferotemporal cortex of the monkey *Science* 166 1303-1306, 1969
GUEDRY F E Psychophysiological studies of vestibular function In *Contributions to Sensory Physiology*, edited by W D Neff New York Academic Press, 1965
GUILD S R Correlations of histologic observations and the acuity of hearing *Acta Otolaryng (Stockholm)* 17 207-249, 1932
GUILLOT M Some characteristics of the phenomenon of partial an osmia *C R Soc Biol (Paris)* 142 161-162 1948a
GUILLOT, M Physiology of the senses Partial anosmia and fundamental odours *C R Acad Sci [D] (Paris)* 226 1307-1309 1948
HABER, R N How we remember what we see *Sci Amer* 222 104-115 1970
HABER, R N, AND M HERSHENSON The effects of repeated brief exposures on the growth of a percept *J Exp Psychol* 69 40-46 1965
HABER, R N, AND L S NATHANSON Processing of sequentially presented letters *Percept Psychophys* 5 359-361 1969
HABER, R N, AND L G STANDING Direct measures of short term visual storage *Quart J Exp Psychol* 21 43-54, 1969
HABER, R N, AND L STANDING Direct estimates of apparent duration of a flash followed by visual noise *Canad J Psychol* 24 216-229, 1970
HAGINS, W A Flash photolysis of rhodopsin in the retina *Nature (London)* 177 989-990, 1956
HAHN, H Über die Ursache der Geschmacksempfindung *Klin Wschr* 15 933-935, 1936
HARDY, J D, H G WOLFF AND H GOODELL *Pain Sensations and Reactions* Baltimore Williams & Wilkins, 1952
HARRIS, G G L S FRISHKOPF, AND A FLOCK Receptor potentials from hair cells of the lateral line *Science* 167 76-79 1970
HARTLINE, H K The response of single optic nerve fibers of the vertebrate eye to illumination of the retina *Amer J Physiol* 121 400-415 1938
HARTLINE, H K, AND C H GRAHAM Nerve impulses from single receptors in the eye *J Cell Physiol* 1 277-295 1932
HAYHOW, W R The cytoarchitecture of the lateral geniculate body in the cat in relation to the distribution of crossed and uncrossed optic fibers *J Comp Neurol* 110 1-64, 1958
HECHT, S Rods, cones and the chemical basis of vision *Physiol Rev* 17 239-290, 1937
HECHT, S S SHLAER, AND M H PIRENNE Energy, quanta and vision *J Gen Physiol* 25 819-840, 1942
HECHT, S, AND C D VERRIJP Intermittent stimulation by light III The relation between intensity and critical fusion frequency for different retinal locations *J Gen Physiol* 17 251-265 1933
HENNING, H Die Qualitatenreike des Geshmacks *Z Psychol* 74 203-219 1916
HENNING H Smell a handbook for the spheres of psychology physiology zoology botany chemistry physics neurology ethnology language literature aesthetics and history Leipzig Verlag 1924 p 453-458
HENSEL H Classes of receptor units predominantly related to thermal stimuli In *Touch Heat and Pain* Ciba Foundation Symposium Boston Little Brown, 1966
HERING E *Beitrage zur Physiologie* Leipzig Engelmann 1861
HERING E *Beitrage zur Physiologie* Leipzig Engelmann 1864
HERNANDEZ PEÓN R C GUZMAN FLORES M ALCAREZ AND A FERNANDEZ GUARDIOLA Sensory transmission in visual pathway during attention in unanesthetized cats *Acta Neurol Lat Amer* 3 1-8 1957
HERRICK R M Foveal luminance discrimination as a function of the duration of the decrement or increment in luminance *J Comp Physiol Psychol* 49 437-443 1956
HERRMAN H, AND F H HICKMAN Exploratory studies on cortical metabolism *Johns Hopkins Med J* 82 225-250 1948
HIND J E Unit activity in the auditory cortex In *Neural Mechanisms of the Auditory and Vestibular Systems* edited by G L Rasmussen and W F Windle Springfield Ill Thomas 1960
HODGKIN A L The subthreshold potentials in a crustacean nerve fiber *Proc Roy Soc London [Biol] Ser B* 126 87-121 1938
HOLMES, G The organization of the visual cortex in man *Proc Roy Soc (London) [Biol] Ser B* 132 348-361 1945
HOLWAY, A H AND L M HURVICH Differential gustatory sensitivity to salt *Amer J Psychol* 49 37-48 1937
HÖNIG, D P Zur Psychophysik des Geschmacksinnes *Philos Stud (Wundt)* 17 576-623 1901
HORN, G, AND R M HILL Modifications of receptive fields of cells in the visual cortex occurring spontaneously and associated with bodily tilt *Nature (London)* 221 186-188, 1969
HOWARD, I P, AND W B TEMPLETON *Human Spatial Orientation* New York Wiley, 1966, p 533
HSIA, Y, AND G H GRAHAM Spectral luminosity curves of protan opic deuteranopic, and normal subjects *Proc Nat Acad Sci USA* 43 1011-1019 1957
HSIA Y, AND C H GRAHAM Color blindness In *Vision and Visual Perception*, edited by C H Graham New York Wiley 1965
HUBBARD R AND A KROPF The action of light on rhodopsin *Proc Nat Acad Sci USA* 44 130-139 1958
HUBEL D H Single unit activity in lateral geniculate body and optic tract of unrestrained cats *J Physiol (London)* 150 91-104 1960
HUBEL, D H AND T N WEISEL Integrative action in the cat's lateral geniculate body *J Physiol (London)* 155 385-398 1961
HUBEL, D H AND T N WIESEL Receptive fields binocular interaction and functional architecture in the cat's visual cortex *J Physiol (London)* 160 106-154, 1962
HUBEL, D H, AND T N WIESEL Receptive fields of cells in striate cortex of very young, visually inexperienced kittens *J Neurophysiol* 26 994-1002, 1963
HUBEL, D H, AND T N WIESEL Receptive fields and functional architecture in two non striate visual areas (18 and 19) of the cat *J Neurophysiol* 28 229-289, 1965
HUBEL, D H AND T N WIESEL Receptive fields and functional architecture of monkey striate cortex *J Physiol (London)* 195 215-243 1968
HUBEL, D H, AND T N WIESEL Cells sensitive to binocular depth in area 18 of the macaque monkey *Nature (London)* 225 41-42, 1970
HULL, C D J GARCIA, N A BUCHWALD B DUBROVSKY, AND B H FEDER Role of the olfactory system in arousal to X ray *Nature (London)* 205 627-628, 1965
IGGO, A Gastric mucosal chemoreceptors with vagal afferent fibres in the cat *Quart J Exp Physiol* 42 398-409, 1957
IGGO, A Cutaneous receptors with a high sensitivity to mechanical displacement In *Touch, Heat and Pain* Ciba Foundation Symposium, Boston Little, Brown, 1966
JERGER, J *Modern Developments in Audiology* New York Academic Press 1963

JOHNSON W H AND V B G TAYLOR Some experiments on the relative effectiveness of various types of accelerations on motion sickness *Aerospace Med* 32 205–208 1961
JOHNSTON J W Infrared loss theory of olfaction untenable *Physiol Zool* 26 266–271 1953
JUDD D B Color perceptions of deuteranopic and protanopic observers *J Opt Soc Amer* 39 252–256 1948
JUDD D B Color perceptions of deuteranopic and protanopic observers *J Res Natn Bur Stand* 4 247–271 1948
JULESZ B Binocular depth perception without familiarity cues *Science* 145 356–362 1964
KALMUS H The familial distribution of congenital tritanopia *Ann Hum Genet* 20 39–56 1955
KATSUKI Y Neural mechanism of auditory sensation in cats In *Sensory Communication* edited by W A Rosenblith Cambridge M I T Press 1961 p 561–583
KATSUKI Y N SUGA AND Y KANNO Neural mechanism of the peripheral and central auditory system in monkeys *J Acoust Soc Amer* 34 1396–1410 1962
KATSUKI Y T SUMI H UCHIYAMA AND T WATANABE Electric responses of auditory neurons in cat to sound stimulation *J Neurophysiol* 21 569–588 1958
KEELE C A Measurement of responses to chemically induced pain In *Touch Heat and Pain* Ciba Foundation Symposium Boston Little Brown 1966
KELLOGG R S AND A GRAYBIEL Lack of response to thermal stimulation of the semicircular canals in the weightless phase of parabolic flight *Aerospace Med* 38 487–490 1967
KELLY D H Visual responses to time dependent stimuli I Amplitude sensitivity measurements *J Opt Soc Amer* 51 422–429 1961
KENSHALO D R Behavioral and electrophysiological responses of cats to thermal stimuli In *The Skin Senses* Springfield Ill Thomas 1968
KERR D I B AND K E HAGBARTH An investigation of olfactory centrifugal fiber system *J Neurophysiol* 18 362–374 1955
KIANG N Y S *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve* Cambridge M I T Press 1966
KIESOW F Contribution a la psycho physiologie de la cavite buccale *Arch Ital Biol* 30 377–398 1898
KIMELDORF D J AND E L HUNT Ionizing radiation neural function and behavior New York Academic Press 1965 p 331
KIMURA D Some effects of temporal lobe damage on auditory perception *Canad J Psychol* 15 156–165 1961a
KIMURA D Cerebral dominance and the perception of verbal stimuli *Canad J Psychol* 15 166–171 1961b
KIMURA, D Left right differences in the perception of melodies *Quart J Exp Psychol* 16 355–368 1964
KIMURA K AND L M BEIDLER Microelectrode study of taste receptors of rat and hamster *J Cell Physiol* 58 131–140 1961
KISTIAKOWSKI G B On the theory of odours *Science* 112 154–155 1950
KOMURO K Le minimum perceptible de l'odorat dans une enceinte absolument inodore *Arch Neerl Physiol* 6 20–24 1921
KONISHI, T AND T YASUNO Summating potential of the cochlea in the guinea pig *J Acoust Soc Amer* 35 1448–1452 1963
KOPERA A Untersuchungen uber die Unterschiedsempfindlichkeit im Bereiche des Geschmacksinns *Arch Ges Psychol* 82 273–307, 1931
KRAKAUER D AND K M DALLENBACH Gustatory adaptation to sweet sour and bitter *Amer J Psychol* 49 469–475 1937
KRUGER L A N FELDZAMEN, AND W R MILES A scale for measuring supra threshold olfactory intensity *Amer J Psychol* 68 117–123 1955a
KRUGER L A N FELDZAMEN AND W R MILES Comparative olfactory intensities of the aliphatic alcohols in man *Amer J Psychol* 68 386–395 1955b
KUFFLER, S W Neurons in the retina Organization, inhibition and excitation problems *Cold Spring Harb Symp Quant Biol* 17 281–292, 1952
KUFFLER, S W Discharge patterns and functional organization of mammalian retina *J Neurophysiol* 16 37–68, 1953
LEE, T S Physiological gustatory sweating in a warm climate *J Physiol (London)* 124 528–542, 1954
LE GRAND *Light, Colour and Vision* London Chapman 1968
LELE P P AND G WEDDELL The relationship between neurohistology and corneal sensibility *Brain* 79 119–154 1956
LE MAGNEN J Analysis of complex and homologous odours by fatigue *C Acad Sci [D] (Paris)* 226 753–754 1948
LETTVIN J Y AND R C GESTELAND Speculations on smell *Cold Spring Harb Symp Quant Biol* 30 217–225 1965
LETTVIN J Y H R MATURANA W S McCULLOCH AND W H PITTS What the frog's eye tells the frog's brain *Proc Instr Radio Engrs* 47 1940–1951 1959
LEVINSON J Z Nonlinear and spatial effects in the perception of flicker In *Flicker* edited by H E Herkes and L H van der Tweel The Hague Junk 1964 p 36–55
LEVINSON J Z Flicker fusion phenomena *Science* 160 21–28 1968
LEWIS T AND W HESS Pain derived from the skin and the mechanism of its production *Clin Sci* 1 39–61 1933
LICKLIDER J C R Basic correlates of the auditory stimulus In *Handbook of Experimental Psychology* edited by S S Stevens New York Wiley 1951 p 985–1039
LÖEWENSTEIN W R Chemical or physical nature of transduction In *Touch, Heat and Pain* edited by A V S de Reuck and J Knight Boston Little Brown 1966 p 137
LOWENSTEIN O AND T D M ROBERTS The equilibrium function of the otolith organs of the thornback ray (Raja clavata) *J Physiol (London)* 110 392–415 1949
LÖWENSTEIN O AND T D M ROBERTS The localization and analysis of the responses to vibration from the isolated elasmobranch labyrinth *J Physiol (London)* 114 471–489 1951
LOWENSTEIN O AND A SAND The activity of the horizontal semicircular canal of the dogfish *Scyllium canicula J Exp Biol* 13 416–428 1936
LOWENSTEIN, O AND A SAND The mechanism of the semicircular canal A study of the responses of single fibre preparations to angular accelerations and to rotation at constant speed *Proc Roy Soc (London) [Biol] Ser B* 129 256 1940
LOWENSTEIN O AND J WERSALL A fundamental interpretation of the electron microscopic structure of the sensory hairs in the cristae of the elasmobranch Raja Clavata in terms of directional sensitivity *Nature (London)* 184 1807–1808 1959
LURIA A R The functional organization of the brain *Sci Amer* 222 66–78 1970
MacNICHOL E F JR AND G SVAETICHIN Electric responses from the isolated retinas of fishes *Amer J Ophthal* 46 26–40 1958
MAGOUN H, AND S RANSON The afferent path of the light reflex A review of the literature *Arch Ophthal (Chicago)* 13 863–874 1935a
MAGOUN, H AND S RANSON The central path of the light reflex *Arch Ophthal (Chicago)* 13 791–811, 1935b
MAKOUS W S NORD B OAKLEY AND C PFAFFMAN The gustatory relay in the medulla In *Olfaction and Taste*, edited by Y Zotterman New York Macmillan 1963, p 381–393
MARKS W B W H DOBELLE AND E F MacNICHOL Visual pigments of single primate cones *Science* 143 1181–1183, 1964
MARSHALL, W H, AND S A TALBOT Recent evidence for neural mechanisms in vision leading to a general theory of sensory acuity *Biol Symp* 7 117–164 1942
McEWEN W K The yellow pigment of human lenses *Amer J Ophthal* 47 144–146 1959
McKINLEY, R W (ed) *IES Lighting Handbook* New York Illuminating Engineering Society 1947
McNEW, J Role of the red nucleus in visually guided behavior of the rat *J Comp Physiol Psychol* 65 282–289 1968
MEIRY J L The vestibular system and human dynamic space orientation Nat Aeronautics Space Admin NASA Cr 628 (Doctoral thesis M I T June 1965) 1966
MELLONI, B J *The Internal Ear, An Atlas of Some Pathological Conditions of the Eye, Ear and Throat* Chicago Abbott Laboratories, 1957 p 26–31
MELZACK, R AND K L CASEY Sensory, motivational and central control determinants of pain In *The Skin Senses* Springfield, Ill Thomas, 1968
MILES, F A Centrifugal effects in the avian retina *Science* 170 992–995, 1970
MILLER E F, AND A GRAYBIEL Otolith function as measured by

ocular counterrolling In *The Role of the Vestibular Organs in Space Exploration* Nat Aeronautics Space Admin NASA SP 77 121-130 1965

MILLER E F JR, AND A GRAYBIEL A role of the otolith organs in the perception of horizontality *Amer J Psychol* 79 24-37 1966

MILLER E F JR AND A GRAYBIEL Effects of drugs on ocular counterrolling In *Third Symposium on the Role of the Vestibular Organs in Space Exploration* Nat Aeronautics Space Admin SP 152 341-349 1968

MILLER, E F JR A GRAYBIEL AND R S KELLOG Otolith organ activity within earth standard one half standard and zero gravity environments *Aerospace Med* 37 399-403 1966

MILLER, J W AND J E GOODSON Motion sickness in a heli copter simulator *Aerospace Med* 31 204-212 1961

MONCRIEFF, R W *The Chemical Senses* New York Wiley 1951

MONEY, K E Discussion of Influence of Linear and Angular Accel erations on Nystagmus by F E GUEDRY JR In *Second Sympo sium on The Role of the Vestibular Organs in Space Exploration* Nat Aeronautics Space Admin , NASA SP 115 196-197 1966

MOTOKOWA K N TAIRA AND J OKUDA Spectral responses of single units in the primate visual cortex *Tohoku J Exp Med* 78 320-337 1962

MOULTON, D G Electrical activity in the olfactory system of rab bits with indwelling electrodes In *Olfaction and Taste* New York Macmillan, 1963

MOULTON, D G Differential sensitivity to odors *Cold Spring Harb Symp Quant Biol* 30 201-206, 1965

MOULTON, D G AND L M BEIDLER Structure and function in the peripheral olfactory system *Physiol Rev* 47 1-52 1967

MOUNTCASTLE V B Some functional properties of the somatic afferent system In *Sensory Communication*, edited by W A Ro senblith Cambridge M I T Press, 1961, p 403-436

MOUNTCASTLE V B Chemical or physical nature of transduction In *Touch, Heat and Pain* edited by A V S de Reuck and J Knight Boston Little Brown 1966 p 137

MOUNTCASTLE, V B W H TALBOT AND H H KORNHUBER The neural transformation of mechanical stimuli delivered to the monkey's hand In *Touch, Heat and Pain*, edited by A V S de Reuck and J Knight Boston Little, Brown, 1966, p 325-345

MOZELL, M M Electrophysiology of the olfactory bulb *J Neuro physiol* 21 183-196, 1958

MOZELL, M M The spatiotemporal analysis of odorants at the level of the olfactory receptor sheet *J Gen Physiol* 50 25-41, 1966

MOZELL, M M , AND C PFAFFMANN The afferent neural process in odor perception *Ann N Y Acad Sci* 58 96-108, 1954

MULLER G E Uber das Aubertasche Phanomen *Z Psychol Physiol Sinnesorg* 49 109-244, 1916

MULLER, J *Beitrage zur vergleichenden Physiologie des Gesicht sinnes* Leipzig Cnoblock, 1826, p 46

NEFF, W D Neural mechanisms of auditory discrimination In *Sen sory Communication*, edited by W A Rosenblith Cambridge M I T Press, 1961

NEISSER, U *Cognitive Psychology* New York Appleton, 1967

NEUHAUS W Odour thresholds of dogs for ionone and ethyl mer captan and their relation to other odour thresholds of dog and man *Z Naturforsch* [B] 9 560-567, 1954

NEWSOM B D , J F BRADY, W A SHAFER AND R S FRENCH Adaptation to prolonged exposures in the revolving space station stimulator *Aerospace Med* 37 778-783, 1966

NOMOTO, M N SUGA, AND Y KATSUKI Discharge pattern and inhibition of primary auditory nerve fibers in the monkey *J Neuro physiol* 27 768-787, 1964

OGAWA, H M SATO, AND S YAMASHITA Multiple sensitivity of chorda tympani fibres of the rat and hamster to gustatory and thermal stimuli *J Physiol (London)* 199 223-240, 1968

OGLE, K N *Researches in Binocular Vision* Philadelphia Saunders, 1950

OHRBACK, J Disturbances of the maze habit following occipital cortex removals in blind monkeys *A M A Arch Neurol Psychol* 81 49-54, 1959

O'KEEFE, J AND H BOUMA Complex sensory properties of certain amygdala units in the freely moving cat *Exp Neurol* 23 384-398, 1969

OLMSTEAD, J M D The role of the autonomic nervous system in accommodation for far and near vision *J Nerv Ment Dis* 99 794-798 1944

OPTICAL SOCIETY OF AMERICA *The Science of Color* New York Crowell, 1953

ØSTERBERG G A Topography of the layer of rods and cones in the human retina *Acta Ophthal* 13 (Suppl VI) 103 pp 1935

OTTOSON D Analysis of the electrical activity of the olfactory epi thelium *Acta Physiol Scand* 35 (Suppl 122) 1-83 1956

OTTOSON D Generation and transmission of signals in the olfactory system In *Olfaction and Taste* edited by Y Zotterman New York Macmillian 1963 p 35-44

OTTOSON D AND G M SHEPHERD Experiments and concepts in olfactory physiology *Progr Brain Res* 23 83-138 1967

PASIK P AND T PASIK Oculomotor functions in monkeys with le sions of the cerebrum and superior colliculi In *The Oculomotor System* edited by M Bender New York Harper 1964

PATTON H D , T C RUCH, AND A WALKER Experimental hypo geusia from Horsley-Clarke lesions of the thalamus *J Neurophysiol* 7 171-184 1944

PFAFFMANN, C Gustatory afferent impulses *J Cell Physiol* 17 243-258 1941

PFAFFMANN C Taste and smell In *Handbook of Experimental Psychology* edited by S S Stevens New York Wiley 1951 chap 29 p 1143-1171

PFAFFMANN, C Gustatory nerve impulses in rat cat and rabbit *J Neurophysiol* 18 429-440 1955

PFAFFMANN C Taste stimulation and preference behavior In *Ol faction and Taste Proc 1st Int Symp , Wenner Gren Center, Stockholm* edited by Y Zotterman New York MacMillan 1963

PFAFFMANN, C , R P ERICKSON, G P FROMMER, AND B P HALPERN Gustatory discharges in rat medulla and thalamus In *Sensory Communication*, edited by W A Rosenblith New York Wiley 1961

PIERON, H Vision in intermittent light In *Contributions to Sensory Physiology* edited by W D Neff New York Academic Press, 1965, vol I 179-264

PIRENNE, M H Spectral luminous efficiency of radiation In *The Eye*, edited by H Davson New York Academic Press, 1962, vol 2 chap 4, p 79

PITT, F H G The nature of normal trichromatic and dichromatic vision *Proc Roy Soc (London) [Biol] Ser B* 132 101-117 1944

POLYAK, S *The Retina* Chicago Univ Chicago Press 1941

POLYAK S *The Vertebrate Visual System* Chicago Univ Chicago Press 1957

PORTMANN G *Diseases of the Ear, Nose, and Throat* Baltimore Williams & Wilkins, 1951

QUILLIAM, T A Unit Design and array patterns in receptor organs In *Touch, Heat and Pain* edited by A V S deReuck and J Knight Boston Little, Brown 1966, p 86-112

RALL, W G M SHEPHERD, T S REESE, AND M W BRIGHTMAN Dendodendritic synaptic pathway for inhibition in the olfactory bulb *Exp Neurol* 14 44-56, 1966

RASHBASS, C The relation between saccadic and smooth tracking eye movements *J Physiol (London)* 159 326-338, 1961

RASHBASS, C , AND G F M RUSSELL Action of a barbiturate drug (Amylo barbitone sodium) on the vestibulo ocular reflex *Brain* 84 329, 1961

RASHBASS, C AND G WESTHEIMER Disjunctive eye movements *J Physiol (London)* 159 339-360, 1961

RASMUSSEN, G L Efferent fibers of the cochlear nerve and coch lear nucleus In *Neural Mechanisms of the Auditory and Vestibular Systems*, edited by G L Rasmussen and W F Windle Springfield Ill Thomas 1960

RUSMUSSEN, A T Studies of eighth cranial nerve of man *Laryngo scope* 50 67-83 1940

RATOOSH, P On Interposition as a cue for the perception of dis tance *Proc Nat Acad Sci U S A* 35 357-359, 1949

REICHARDT, W , AND G MACGINITIE Zur Theorie der lateralen Inhibition *Kybernetik* 1 155-165, 1962

RETZIUS, G *Das Gehororgan der Wirbelthiere* Stockholm Samson and Wallin 1884

RIESEN, A H Plasticity of behavior psychological aspects In *Biological and Biochemical Bases of Behavior*, edited by H F Harlow and C N Woolsey Madison Univ Wisconsin Press, 1958

RIGGS L A The human electroretinogram A M A *Arch Ophthal* 60 759–754 1958

RIGGS L A Light as a stimulus for vision In *Vision and Visual Perception* edited by C H Graham New York Wiley 1965a

RIGGS L A Visual acuity In *Vision and Visual Perception* edited by C H Graham New York Wiley 1965b

ROBERTS I D M *Neurophysiology of Postural Mechanisms* New York Plenum 1967 p 354

ROBINSON D A Eye movement control in primates *Science* 161 1219–1224 1968

ROBINSON D A Oculomotor unit behavior in the monkey *J Neurophysiol* 33 393–404 1970

ROBINSON D A A D A F FUCHS Eye movements evoked by stimulation of frontal eye fields *J Neurophysiol* 32 537–648 1969

ROSENBERG B T N MISRA AND R SWITZEER Mechanism of olfactory transduction *Nature (London)* 217 423–427 1968

ROSENZWEIG M R Representations of the two ears at the auditory cortex *Amer J Physiol* 167 147–158 1951

ROSS D A Electrical studies on the frog's labyrinth *J Physiol (London)* 86 117 1936

ROSSI G *Le Basi Morfofunzionali del Controllo delle Sensazioni Acustische* edited by L Candiollo G Filogamo and G Rossi Torino Italy Societa Italiana di Laringologia Otologia e Renologia 1965

ROTHMAN S The nature of itching *Res Publs Ass Nerv Ment Dis* 23 110–122 1943

ROWBOTHAM G F Observations on the effects of trigeminal denervation *Brain* 62 364–380 1939

RUPPERT A G MOUSHEGIAN AND R GALAMBOS Unit responses to sound from auditory nerve of the cat *J Neurophysiol* 26 449–465 1963

RUSHTON W A H *Visual Pigments in man* Sherrington Lectures Springfield Ill Thomas 1962 vol 6 38 pp

RUSHTON W A H Interpretation of retinal densitometry *J Opt Soc Amer* 54 273 1964

SACHS M B AND N Y KIANG Two tone inhibition in auditory nerve fibers *J Acoust Soc Amer* 43 1120 1968

SALA O Some remarks on the vestibular efferent system In *Structure and Function of Inhibitory Neural Mechanisms* edited by C VonEuler S Skoglund and V Soderberg New York Pergamon 1968 p 563

SANDERSON K J I DARIAN SMITH AND P O BISHOP Binocular corresponding receptive fields of single units in the cat dorsal lateral geniculate nucleus *Vision Res* 9 1297–1303 1969

SATO M The effect of temperature change on the response of taste receptors In *Olfaction and Taste* edited by Y Zotterman New York Macmillan 1963

SATO M Gustatory response as a temperature dependent process In *Contributions to Sensory Physiology* edited by W D Neff New York Academic Press 1967

SAWA M AND J DELGADO Amygdala unit activity in the unrestrained cat *Electroenceph Clin Neurophysiol* 15 637–650 1963

SCHADE O H Optical and photoelectric analog of the eye *J Opt Soc Amer* 46 721–739 1956

SCHILLER P H The discharge characteristics of single units in the oculomotor and abducens nuclei of the unanesthetized monkey *Exp Brain Res* 10 347–362 1970

SCHUKNECHT H F (ed) *Otosclerosis* Boston Little Brown 1962

SCHWARZ W Electron microscopic observations of the human vitreous body In *The Structure of the Eye* edited by G K Smelser New York Academic Press 1961

SECHZER J A Successful interocular transfer of pattern discrimination in split brain cats with shock avoidance motivation *J Comp Physiol Psychol* 58 76–83 1964

SEN GUPTA P Olfactory receptor reaction to the lesion of the olfactory bulb In *Olfaction and Taste* edited by T Hayashi New York Macmillan 1967 vol II

SETTERFIELD, W, R G SCHOTT AND L H SNYDER Studies in human inheritance XV The bimodality of the threshold curve for the taste of phenylthio carbamide *Ohio J Sci* 36 231–235 1936

SHANKWEILER, D AND M STUDDERT KENNEDY Identification of consonants and vowels presented to left and right ears *Quart J Exp Psychol* 19 59–63 1967

SHERRINGTON C S *The Integrative Action of the Nervous System* London Constable, 1906

SHLAER S The relation between visual acuity and illumination *J Gen Physiol* 21 165–188 1937

SHORF I E A contribution to our knowledge of taste sensations *J Physiol (London)* 15 191–217 1892

SIMMONS F B Monaural processing In *Foundations of Modern Auditory Theory* edited by J V Tobias New York Academic Press 1970 vol I p 345–379

SJOBERG A Experimental and clinical experiences and comments on ultrasonic treatment of Meniere's disease In *Fourth Symposium on the Role of the Vestibular Organs in Space Exploration* Nat Aeronautics Space Admin NASA SP 187 271 281 1965

SLOAN L I Congenital achromatopsia a report of 19 cases *J Opt Soc Amer* 44 117–128 1954

SMITH C A AND G I RASMUSSEN Recent observations on the olivocochlear bundle *Ann Otol* 72 489–506 1963

SMITH C A AND G L RASMUSSEN Nerve endings in the maculae and cristae of the chinchilla vestibule with a special reference to the efferents In *Third Symposium on the Role of the Vestibular Organs in Space Exploration* Nat Aeronautics Space Admin NASA SP 152 183–200 1967

SPERLING G The information available in brief visual presentations *Psychol Monog* 74 No 11 (Whole No 498) 1960

SPERLING G Successive approximations to a model for short term memory *Acta Psychol* 27 285–292 1967

SPERRY R W Cerebral organization and behavior *Science* 133 1749–1757 1961

SPINELLI D W AND T W BARRETT Visual receptive field organization of single units in the cat's visual cortex *Exp Neurol* 24 76–98 1969

SPOENDLIN H H Ultrastructural studies of the labyrinth in squirrel monkeys In *The Role of the Vestibular Organs in the Exploration of Space* Nat Aeronautics Space Admin NASA SP 77 7–22 1965

STEINHAUSEN W Ueber die Beobachtung der Arpula in den Bogengangsampullen des Labyrinthes des lebenden Hechts *Pflueger Arch* 232 500–512 1933

STERN C *Human Genetics* Ed 2 San Francisco Freeman 1960

STEVENS S S *Handbook of Experimental Psychology* New York Wiley 1951 chap 29 p 1143–1171

STEVENS S S Some similarities between hearing and seeing *Laryngoscope* 68 508–527 1958

STEVENS S S AND H DAVIS *Hearing Its Psychology and Physiology* New York Wiley 1938 p 345

STEVENS S S H DAVIS AND M H LURIE The localization of pitch perception on the basilar membrane *J Gen Psychol* 13 297–315 1934

STEVENS S S AND E H GALANTER Ratio scales and category scales for a dozen perceptual continua *J Exp Psychol* 54 377 1957

STILES W S AND B H CRAWFORD The luminous efficiency of rays entering the eye pupil at different points *Proc Roy Soc (London) [Biol] Ser B* 112 428–450 1933

SVAETICHIN G AND E F MACNICHOL JR Retinal mechanisms for chromatic and achromatic vision *Ann N Y Acad Sci* 74 385–404 1958

TASAKI I Nerve impulses in individual auditory nerve fibers of guinea pig *J Neurophysiol* 17 97–122 1954

TASAKI I H DAVIS AND D H ELDREDGE Exploration of cochlear potentials in guinea pig with a microelectrode *J Acoust Soc Amer* 26 765–773 1954

TASAKI I H DAVIS AND J P LE GOUIX The space time pattern of the cochlear microphonics (guinea pig) as recorded by differential electrodes *J Acoust Soc Amer* 24 502–519 1952

TASAKI I AND C FERNANDEZ Modifications of cochlear microphonics and action potentials by KCl solution and by direct currents *J Neurophysiol* 15 497–512 1952

TAYLOR H L J C SMITH A H WALL AND B CHUDDOCK Role of the olfactory system in the detection of X rays by the rhesus monkey *Physiol Behav* 3 929–933, 1968

TEAS D C Cochlear Processes In *Foundations of Modern Auditory Theory*, edited by J V Tobias New York Academic Press, 1970, vol I p 255–303

TEITELBAUM, P AND A N EPSTEIN The role of taste and smell in the regulation of food and water intake In *Olfaction and Taste*

Proc 1st Int Symp Wenner Gren Center, edited by Y Zotterman New York Macmillan 1963

TITCHEN D A Reflex stimulation and inhibition of reticulum contractions in the ruminant stomach *J Physiol (London)* 141 1-21 1958

TOBIAS J V (ed) *Foundations of Modern Auditory Theory* New York Academic Press 1970 vol I

TONNDORF J Dimensional analysis of cochlear models *J Acoust Soc Amer* 32 493-497 1960

TSCHERMAK SEYSENEGG A Über Parallaktoskopie *Pflueger Arch Physiol* 241 454-469 1939

TUCKER D Olfactory cilia are not required for receptor function *Fed Proc* 26 544 1967

TUCKER, D, AND L M BEIDLER Autonomic nervous system influence on olfactory receptors *Amer J Physiol* 187 637 1956

VERRILLO, R T A duplex mechanism of mechanoreception In *The Skin Senses*, edited by D R Kenshalo Springfield Ill Thomas 1968

VON BÉKÉSY G Über die Resonanzkurve und die Ablingzeit der verschiedenen Stellen der Schneckentrennwand *Akust Zh* 8 66-76 1943

VON BÉKÉSY, G The variation of phase along the basilar membrane with sinusoidal vibrations *J Acoust Soc Amer* 19 452-460 1947

VON BÉKÉSY G D C resting potentials inside the cochlear partition *J Acoust Soc Amer* 24 72-76, 1952

VON BÉKÉSY G Shearing microphonics produced by vibrations near the inner and outer hair cells *J Acoust Soc Amer* 25 786-790 1953

VON BÉKÉSY G *Experiments in Hearing* New York McGraw-Hill 1960

VON BÉKÉSY, G Olfactory analogue to directional hearing *J Appl Physiol* 19 369-373 1964a

VON BÉKÉSY, G Sweetness produced electrically on the tongue and its relation to taste theories *J Appl Physiol* 19 1105-1113, 1964b

VON BÉKÉSY G Taste theories and the chemical stimulation of single papillae *J Appl Physiol* 21 (Suppl 1) 1-9, 1966

VON FREY M Über die zur ebenmerklichen Erregung des Drucksinns erfordlichen Energiemengen *Z Biol* 70 333-347, 1919

VON HELMHOLTZ H L F *Treatise on Physiological Optics* translated by J P C Southall Rochester, N Y Opt Soc Amer 1924

VON SKRAMLIK, E The fundamental substrates of taste In *Olfaction and Taste*, edited by T Hayashi New York Macmillan, 1963

WAALER, G H M Ueber die Erblichkeitsverhaltnisse der verschiedenen Arten von augeborenen Rotgrunblinkheit *Z Indukt Abstamm u VererbLehre* 45 279-333, 1927

WALD G, AND P K BROWN Human rhodopsin *Science* 127 222-226 1958

WALD G, AND P K BROWN Human color vision and color blindness *Cold Spring Harb Symp Quant Biol* 30 345-359 1965

WALD G P K BROWN, AND I R GIBBONS The problem of visual excitation *J Opt Soc Amer* 53 20-35, 1963

WALD, G AND A B CLARK Visual adaptation and the chemistry of the rods *J Gen Physiol* 21 93-105, 1937

WALLS G L *The Vertebrate Eye* Bloomfield Hills Mich Cranbrook Institute of Science, 1942

WEALE, R A Cone monochromatism *J Physiol (London)* 121 548-569, 1953

WEALE, R A Light absorption by the lens of the human eye *Opt Acta (London)* 1 107-110, 1954

WENDT G R Vestibular functions In *Handbook of Experimental Psychology* edited by S S Stevens New York Wiley, 1951 p 1191-1223

WERNER, G, AND V B MOUNTCASTLE Quantitative relations between mechanical stimuli to the skin and neural responses evoked by them In *The Skin Senses* Springfield, Ill Thomas, 1968

WERSALL, J Studies on the structure and innervation of the sensory epithelium of the cristae ampullares of the guinea pig *Acta Otolaryng (Stockholm)* (Suppl 126) 1 1956

WERSALL, J Efferent innervation of the inner ear In *Structure and Function of Inhibitory Neuronal Mechanisms* edited by C Euler et al New York Pergamon 1968 p 123-129

WERSALL, J A FLOCK AND P G LUNDQUIST Structural basis for directional sensitivity in cochlear and vestibular sensory receptors *Cold Spring Harb Symp Quant Biol* 30 115-132 1965

WERTHEIM T Über die indirekte Sehscharfe *Z Psychol* 7 172-187 1894

WESTHEIMER, G Modulation thresholds for sinusoidal light distributions on the retina *J Physiol (London)* 152 67-74 1960

WESTHEIMER G AND C RASHBASS Barbiturates and eye vergence *Nature (London)* 191 833-834 1961

WEVER E G AND C W BRAY Action currents in the auditory nerve in response to acoustical stimulation *Proc Nat Acad Sci USA* 16 344-350 1930

WEYMOUTH, F W The eye as an optical instrument In *A Text book of Physiology* Ed 17 edited by J F Fulton Philadelphia Saunders 1955 chap 23

WHITFIELD I C The pharmacological behaviour of the cochlear nucleus In *Drugs and Sensory Functions* edited by A Herxheimer London Churchill 1967

WHITFIELD I C AND E F EVANS Responses of auditory cortical neurons to stimuli of changing frequency *J Neurophysiol* 28 655-672, 1965

WIDÉN L AND C AJMONE MARSON Action of afferent and corticofugal impulses on single elements of the dorsal lateral geniculate nucleus In *Das Visuelle System*, edited by R Jung and H Kornhuber Berlin Springer 1961 p 125-132

WIESEL, T N AND D H HUBEL Comparison of the effects of unilateral and bilateral eye closure on cortical unit responses in kittens *J Neurophysiol* 28 1029-1040, 1965

WIESEL T N, AND D H HUBEL Spatial and chromatic interactions in the lateral geniculate body of the Rhesus monkey *J Neurophysiol* 29 1145 1966

WILSKA, A Eine Methode zur Bestimmung der Horschwellanamplituden des Trommelfels bei verschiedenen Frequenzen *Skand Arch Physiol* 72 161-165, 1935

WINTRINGHAM W T Color television and colorimetry *Proc Inst Radio Engn* 39 1135-1172 1951

WOELLNER R C, AND A GRAYBIEL Counter rolling of the eyes and its dependence on the magnitude of gravitational or inertial force acting laterally on the body *J Appl Physiol* 14 632-634 1959

WOOD, C D Use of drugs in the prevention of motion sickness In *Fourth Symposium on the Role of the Vestibular Organs in Space Exploration* Nat Aeronautics Space Admin NASA SP 187 101-106, 1968

WOODSON W E *Human Engineering Guide for Equipment Designers* Los Angeles Univ California Press, 1954

WRIGHT, R H *The Science of Smell* New York Basic Books, 164 pp, 1964

WRIGHT, R H, J R HUGHES, AND D E HENDRIX Olfactory coding *Nature (London)* 216 404-406 1967

WRIGHT, W D *The Measurement of Colour* New York Macmillan, 1958, p 263

YOUNG, T *Lectures in Natural Philosophy* London Printed for Jos Johnson St Paul s Church Yard, by Wm Savage 1807, vol 2

ZACKS, J L Temporal summation phenomena at threshold their relation to visual mechanisms *Science* 170 197-199, 1970

ZOTTERMAN, Y Special senses Thermal receptors *Ann Rev Physiol* 15, 357-372, 1953

ZWAARDEMAKER, H *L odorat* Paris Doin, 1925

ZWISLOCKI, A Analysis of some auditory characteristics In *Handbook of Mathematical Psychology* edited by R D Luce, R R Bush, and E Galanter New York Wiley, 1965, vol III

9

Neural Control Systems

1 / ANALYSIS OF NEURAL CONTROL SYSTEMS

Classes of Neural Components

Every neural control system is made up of relatively few types of cells The control is always exerted finally upon a muscle (either striated, cardiac, or smooth in its histology) or upon a gland (either exocrine or endocrine) Thus, there are only two general types of *effector cells*, muscle and gland For these two kinds of cells, however, there are three types of control (Fig 9 1) One type is peculiar to skeletal muscle, another is common to cardiac muscle, smooth muscle, exocrine glands, and a certain few of the endocrine glands, the third is reserved for endocrine glands only In each instance the operation of the control system brings about the shortening or lengthening of the muscle, or the secretion or cessation of secretion of a gland

EFFECTORS AND DETECTORS

Nerves that control muscles and glands are known as *effector nerves* Most of them leave the central nervous system via the ventral roots of the spinal cord or the equivalent roots of cranial nerves Consequently, they are known as *efferent* because their course is away from the brain or spinal cord They are also known as motor nerves, although this term is hardly appropriate for those that control glands

In contrast to the effector cells are those specialized cells that respond to local or environmental changes by sending signals into the nervous system These cells are known as receptors, as *detectors*, or as "sensory" cells if this last term is understood not to imply conscious sensation Their nerves are called *afferent* because they run toward spinal cord and brainstem, they are also called "sensory" merely because there is no other convenient word that is more precise in its meaning Whereas the cell bodies of effector nerves are located in the central nervous system or in ganglia of the autonomic nervous system (pp 9-53–9-65), the cell bodies of detector nerves ordinarily lie in ganglia on the dorsal roots of spinal nerves or within the homologous ganglia of the cranial nerves

In general, the principles of organization of detector-effector systems are similar in spinal and in cranial nerves, with two exceptions These two are the first two of the cranial nerves, the olfactory and optic Fibers in these nerves are not processes arising from detector mechanisms Rather, they are the analogues of fiber tracts within the central nervous system This comes about because the olfactory bulb and the retina are derived from the embryonic brain and not from the neural crest, as are the other peripheral and cranial sensory nerves In fact, the fibers of the olfactory and optic "nerves" are longitudinally coursing commissures of the brain

There are a very few types of neural control where the system includes only the detector cell and the detector nerve, and then a direct connection with the effector nerve and thence with the effector cell Such a system is known as a *two-neuron arc* or *monosynaptic reflex* Under laboratory conditions this type of reflex can be demonstrated by stimulation of a dorsal root so as to evoke a reflex discharge in the ventral root Such reflexes are transmitted through the spinal cord with a shorter delay than other types of transmission, and have been employed as measures of the time required for the afferent neuron to excite the efferent neuron of the reflex The site of this excitation is the *synapse*, a region of intimate contact but not fusion of the two neurons Delay in transmission across a single synapse is in the neighborhood of 0 6 msec (Renshaw, 1940) What happens in this interval is de-

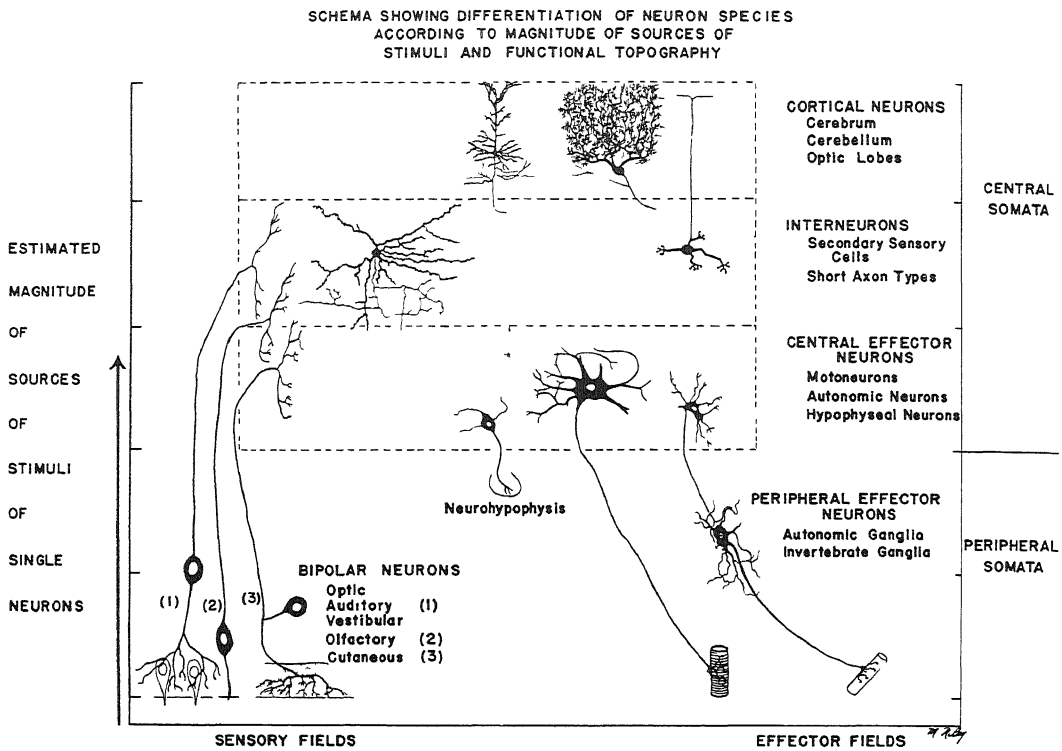


Fig 9 1 Major neuron types in mammalian central nervous system classified according to general function hierarchical level and probable diversity and magnitude of sources of synaptic connections Bodian in Quarton et al 1967

scribed in section 1 of this text, and reconsidered briefly on pp 9-76–9-82 The experimental situation where these measurements are made, however, is scarcely a physiological one, and it seems doubtful that there is any naturally occurring control process where only two neurons can be said to participate in the response One can speak, therefore, of a two-neuron reflex, but probably there is no two-neuron control system for skeletal muscle

CONNECTORS

If two-neuron arcs are not common, then most responses must involve units other than the detector-effector combination These other units are called *connectors*, or internuncial cells or interneurons Their morphology is essentially the same as that of detectors and effectors, in that they receive, propagate, and transmit a disturbance, i e , a *nerve impulse* By their presence in a neural pathway they increase the delay of transmission because of the introduction of their additional synapses Their significance, however, far exceeds this perhaps trivial phenomenon, because they have the highly important property of being able to reverse the "sign" of neural activity If the customary effect of a detector neuron upon an effector neuron is excitation, the interposition of a connector neuron may convert the process into its opposite, namely, a suppression of activity This suppression is known as *inhibition* It is inhibition in the neural control system that permits one muscle to relax when its opposing muscle contracts, and that makes possible the almost infinite gradation of the force of any movement

Another function of connector neurons is the introduction of complexity into control systems (Fig 9 2) A network of nerve fibers linking detectors with effectors via multiple alternate pathways that include connectors has the possibility of actions more complex and also more subtle than any two-neuron system In this sense the connector neurons are analogous to the plug-in cords of a telephone manual switchboard It is true that every telephone could have a direct line to every other telephone in the system But it is apparent that the bulk of wiring to every telephone

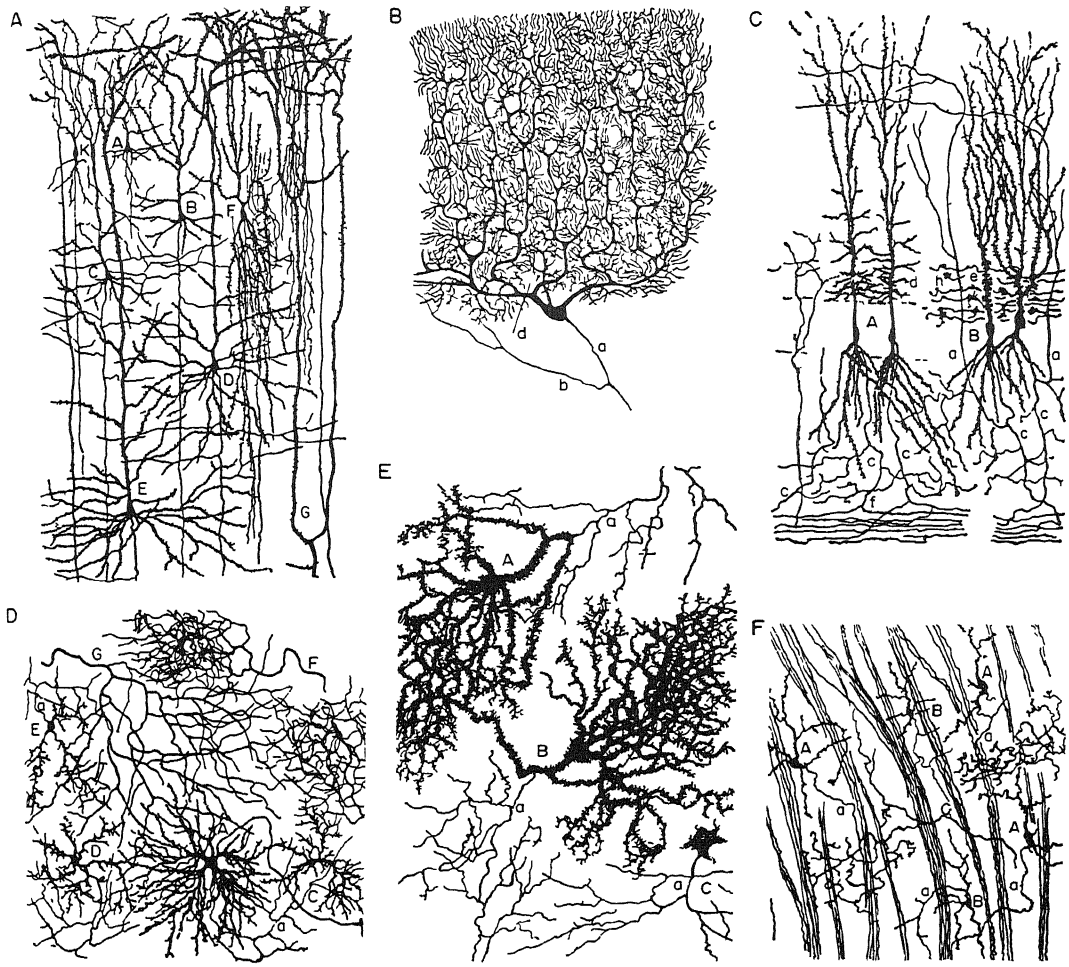


Fig 9 2 Examples of dendritic patterns of neurons which constitute the major cell types of elements in different structures of the mammalian brain *A* outermost three layers of frontal cortex of a 1-month-old human infant Small (*A B C*) and large (*D* and *E*) medium pyramidal neurons as well as the dentritic process of a large pyramidal cell (*G*) of the fourth layer are shown Cells with double dendritic bouquets (*F*) and fusiform appearance (*K*) are also identified *B* Purkinje cell of the cerebellum from adult human Axon (*a*) and axon-collateral (*b*) capillary spaces (*c*) and spaces occupied by basket cells (*d*) are indicated *C,* Pyramidal cells of hippocampus from a 1-month-old rabbit *A* small pyramidal cells of the superior region *B* large pyramidal cells of the inferior region Large ascending collaterals (*a*) axons (*c*) and sites of contact of mossy fibers (*h*) are to be noted *D* Frontal section of the thalamic somatic sensory nucleus of the cat a few days old *A* cell with a long axon *C D* and *E* cells with short axons *F* sensory fibers *G* axons of cortical origin terminating in the sensory nucleus *E* Cells of the pons from a human infant few days old *A* and *B* cells with axons arising from dendrites *C* cell with a bifurcated axon *F* Sagittal section of the caudate nucleus from a newborn rat *A* cells with a long axon *B,* cells with short axon *C* ascending afferent fibers From Ramon ŷ Cajal cited by Purpura 1967

would thereby become enormous By making these interconnections at a central station an economy of number and bulk of peripheral transmission lines is achieved, and the possibility of interconnections is multiplied One sees from this analogy that peripheral nerves can be fine because they only transmit nerve impulses, whereas the spinal cord, brainstem, cerebellum, and cerebral cortex must be relatively more bulky if they are to contain the central stations for the switching processes The size of a given part of the nervous system is largely a function of the number of neurons it contains, and hence is an index of the variety and complexity of its function

In their ability to alter the nature of activity

in a neural circuit, either to increase it—*facilitation*—or to inhibit it, and in bringing into any circuit the possibility of a wealth of alternate connections, the interneurons give the nervous system an enhanced capacity for what Sherrington (1906) called an *integrative action* By this term he meant the creation of an individual from its component reactions Relaxation of an extensor muscle during contraction of a flexor is an integrative action, as is the contraction of extensors in the left leg when the right leg is flexed A certain degree of integration is programmed into even the simplest reflexes in that the detectors make connection with appropriate effectors, rather than with effectors remote and unrelated to the initial stimulus Yet in most instances and especially in higher components of the system the integration depends largely upon interneurons The nature of the individual assembled from unitary responses is a function, therefore, of the number of connector units, the diversity of their connections, and the complexity and variety of controls their activity brings into play [1]

Methods of Physiological Study

ANALYSIS OF NEUROLOGICAL DEFICITS

A "natural" method for studying control systems is provided by disease and trauma affecting the nervous system of human patients, and domestic or other animals Analysis of deficits under such circumstances is the oldest method of study, and the one upon which a large fraction of all that is known about the human brain depends Injuries, known as *lesions*, occur as a result of nutritional deprivation, poisoning, infection, tumors, vascular occlusion, hemorrhage, and both indirect and direct mechanical trauma Because certain parts of the nervous system are relatively small, even a tiny lesion may result rather promptly in a considerable variety of deficits The patient may complain of *symptoms* produced by the lesion, while his physician discovers objective *signs* of abnormal function. It is the spinal cord and brainstem where small injuries may lead to pronounced changes In the cerebral cortex by contrast, where functions are more subtle and the systems seem to possess a high degree of "redundancy," even relatively large lesions may go unrecognized The objectives of a neurological examination are to identify the site of the lesion, to delimit its size, to discover whether it is increasing in size or severity, and if so, how quickly, and to decide if possible what has caused the injury Making these decisions not only requires careful and detailed examination, it also requires a comprehensive knowledge of the location of mechanisms within the neuroaxis In many instances the knowledge that is most useful is the distribution of spinal nerves, and the location of cranial nerve nuclei in the brainstem The former can be ascertained easily from published maps The latter, however, is of importance so great as to be deserving of being memorized by every physician Because the cranial nerve nuclei extend through the brainstem from upper mesencephalon to lower medulla oblongata, when they are included in a lesion the resulting signs or symptoms may give invaluable information about where the lesion is and how extensive it may be

Neurological examination utilizes an ancient method of study in which a patient is either asked to perform some act, is requested to describe some experience, or is stimulated so as to provoke a sensation or a motor response In each of these procedures an *input* to the nervous system is created either verbally or by stimulation, and an *output* is always hoped for either in speech or other motor response In a qualitative sense these techniques analyze the input-output relation of the part of the nervous system being investigated Methods that are in their logic exactly the same are utilized by engineers in analysis of electrical and electronic instruments—or by a mechanic working on a motor car, for that matter Opening the throttle to discover what noise the motor makes is an example of input-output analysis of a crude variety If it is possible to quantitate the stimulus applied, for example, to the skin or to one of the organs of special sensation, and if it is possible to measure the strength of the response, then the techniques of neurophysiology come even closer to those utilized for study of inanimate systems One of the most striking examples of such an experiment

[1] Electrical and mathematical models of how neurons interact may be found in papers by Harmon and Lewis (1966), Kennedy, et al (1969), and George (1970).

Table 9 1
Strength of contraction of *M Tibialis Anticus* (flexor of the foot) in response to stimulation of various cutaneous sensory nerves Upon direct stimulation of the motor nerve the maximal force of contraction of the muscle was 2160 gm

Afferent Nerve Stimulated	Tension of Maximal Reflex Tetanus in Grams	Reflex Tension Expressed as Percentage of Maximal Motor Tetanus
Internal saphenous	800	32
Superficial obturator	165	6 7
Deep obturator	400	16
Nerve to quadriceps and sartorius	1190	44
Musculocutaneous branch of peroneal	1700	69
External plantar	1240	50
Internal plantar	1330	54
Small sciatic	680	28
Hamstring	565	23
Nerve of sural triceps	300 (rather low)	12
Total	8370	

Creed et al , 1932

is found in a table published by Sherrington and his associates (Creed, et al , 1932) and reproduced as Table 9 1, it compares the strength of muscle contraction after stimulation of various cutaneous, sensory nerves Sherrington was able to predict the nature of the connections between afferent and efferent neurons in the spinal cord on the basis of the data in this table In these experiments the input was strong electrical stimulation of the sensory nerves one after the other, the output was the strength of the resulting contraction of the tibialis anticus muscle

Deficits following Experimental Operations. During the development of the specialty of neurosurgery in the early years of this century, surgeons began to operate upon experimental animals for the purpose of learning more about functions of the brain, so that neurological diagnosis and surgical intervention might be more effective At first the experimental lesions were made by ablation, or by incision, and they involved principally the hemispheres of cerebrum or cerebellum because they are accessible to such operations The deficits that follow such an ablation may yield important information about the nature of a control system if some component of the system is well localized in a given part of the brain This is true of the principal motor and sensory systems that are represented just ahead of and just behind the central sulcus on the lateral convexity of the cerebrum But if a given system is diffusely organized, as seems to be the case for other cortical mechanisms, or if a given part of the brain includes components belonging to many different systems, the technique of ablation may have limited usefulness, and may actually yield confusing results The neurosurgical approach is handicapped also by the obvious difficulty of removing an underlying structure, such as one of the basal ganglia, without injury to the overlying tissue

Stereotaxic Surgery. To obviate some of these limitations, and especially to make possible the study of subcortical structures, an English physician-physiologist, Robert Henry Clarke, and his better known neurosurgical colleague, Sir Victor Horsley, invented a technique known as stereotaxic surgery (see Davis, 1964) For this invention Clarke has been called "the Galileo of neurophysiology" (Tepperman, 1970) His method employs a calibrated frame fastened to the head of the subject On the frame is a carrier that can be moved from side to side, from front to back, or up and down, it carries one or more instruments that can be thereby oriented at predetermined positions within the brain If the instrument inserted is an insulated wire, it may be used for stimulation, for destruction of tissue by electrolysis, or for recording (Fig 9 3) Lesions made by electrolysis using direct current may be as small as a millimeter in diameter, or up

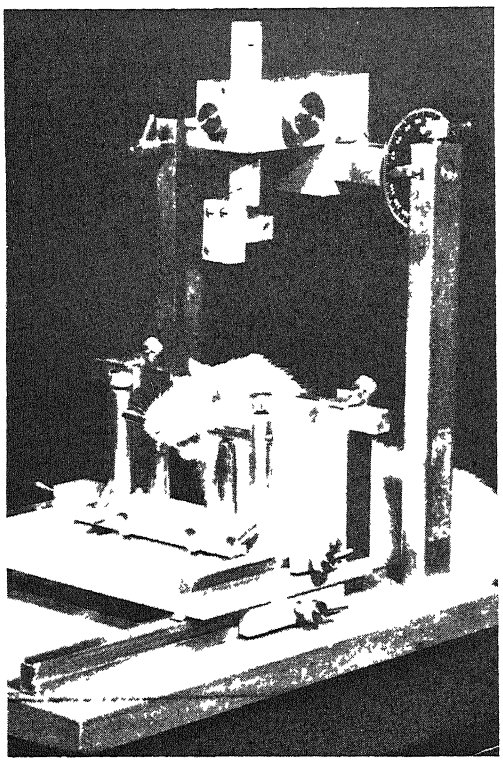


Fig 9 3 Stereotaxic instrument (Baltimore) for use on cat rat or mouse The head holder moves rostrocaudally the electrode carrier moves from side to side and up and down and rotates around the top pivots

to a centimeter if the electrolysis is repeated at many positions (Carpenter and Whittier, 1952) The method may be used also for injection of material through a needle or hollow tube, or for incision of tissue by means of knifelike wires designed for this purpose Stereotaxic instruments are available commercially for animals as small as mice, as well as for operations upon human patients (Fig 9 4). They may be used for *implanting* electrodes or tubes, in an operation performed under anesthesia where the implanted instrument is fixed to the skull so that it may be utilized long after the subject has recovered from the effects of the operation (see Anon , 1971) All of these advantages combine to give stereotaxic operations a pre-eminent place in the techniques of neurophysiology

Recovery from Operations. One of the interesting characteristics of lesions made experimentally, which has a parallel in the experience of many patients with vascular lesions of the cerebrum, is the discrepancy between the initial and the ultimate deficit (Stein et al , 1969) For example, operations upon the rostral hypothalamus of cats may be followed by a fever during which the animal dies, although if the cat is cooled artificially during this period and kept alive until the fever disappears it will attain a fairly normal temperature (Clark et al , 1939) The observation that recovery occurs in patients who have a paralysis, or a sensory disturbance, or a speech defect as a result of occlusion of a cerebral blood vessel is a common clinical experience In many cases the recovery is regarded as "complete," even though neurons of the central nervous system are known to have practically no capacity for regeneration of severed processes Moreover, in mammals, after birth neurons can no longer be produced by multiplication or differentiation of primordial cells How then can recovery be explained?

The first and most obvious possibility is that neurons made inactive by the injury may later recover their function This type of recovery is known to occur after pressure is applied to a nerve, after cooling, anoxia, or the action of local anesthetics Very probably such a se-

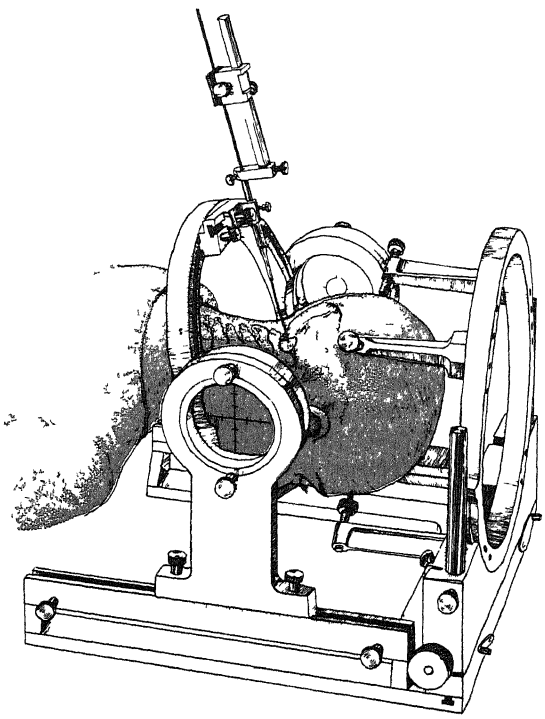


Fig 9 4 Stereotaxic operation upon a human subject The instrument is in place for insertion into the cervical spinal cord Courtesy of Dr E M Todd San Marino California

quence of events takes place at the margins of a lesion following vascular occlusion or electrolysis It may be that the number of temporarily inactivated neurons far exceeds those cells that are in fact "destroyed," with the result that recovery does indeed depend upon restoration of function in many neurons Another possibility is that the lesion has not fatally damaged the whole system, but has removed one component such that remaining parts are temporarily out of balance An automobile may go too fast when the brakes fail suddenly, but it is possible to drive a car at conventional speeds even if it has no brakes With reference to the fever after injury to the rostral hypothalamus, the hyperthermia may be a result of release of the system from a usual hypothalamic inhibition But the fever may disappear when other parts of the system once again achieve a more harmonious balance, even though a component no longer functions Another example may be the phenomenon of recovery from spinal "shock" After transection of the spinal cord there is depression or absence of spinal reflexes below the section, in most laboratory animals these reflexes gradually return, the duration of the recovery ranging from minutes or days in rats, to days and weeks in cats The conventional explanation is that removal of facilitatory influences from higher mechanisms left the spinal neurons at first relatively unexcitable, but that they later recovered the ability to respond to reflex type of stimulation even in the absence of descending facilitation

A third possible explanation is that other neurons have taken over the function of those that were destroyed This possibility is uncertain and controversial (Lashley, 1929), and perhaps related to the phenomenon called "learning" In the spinal cord and brainstem it seems highly questionable whether such a transfer of function can occur, because even tiny but well-placed lesions may cause permanent deficits Finally, there is the possibility that the function being analyzed may be effected through redundant systems, and that the lesion has removed only a part of the participating mechanisms What is called recovery may be simply the action of the "back-up" systems' taking over the function If this is true, it should be possible by suitable testing to demonstrate that a deficit remains in spite of the recovery This can be done in animals that have survived the postoperative fever of rostral hypothalamic injury After they have regained a normal body temperature under laboratory conditions, they experience another bout of fever if they are exposed to a warm environment This observation demonstrates another important point, moreover, that a test is useful only if it succeeds in revealing a specific deficit The input of the test must be selected so that it will disclose whether or not some defect in output is present

Analysis of Subtle Deficiencies. One sees, therefore, that the value of any study of a patient or a laboratory animal depends upon the ingenuity and care with which any deficit is displayed and analyzed In clinical neurology one of the most significant of the tests available is simply careful observation of a patient's posture and locomotion More than a casual approach is needed, however, along with a thorough knowledge of the location of both sensory and motor mechanisms in cerebrum, cerebellum, brainstem, and spinal cord One is inclined to emphasize the value of a systematic neurological examination because so many students suppose that only a neurologist can perform one adequately On the contrary, the fundamentals of a neurological examination should be within the skill of every physician

In the design of neurological tests the discipline of experimental psychology has had an important influence The use of conditioning techniques, those requiring solution of problems such as maze running, the ability to discriminate between minimal differences in sensory cues, and many others have greatly extended the scope of study of neurological deficits (chap 8) It is in this direction that further progress in understanding of the nervous system is likely to take place, as new and imaginative techniques are devised for disclosure of hitherto unknown input-output relationships

ARTIFICIAL STIMULATION

A second type of study of the nervous system, either intact or following some operation or injury, depends upon the use of a specific, controlled stimulus, of measured intensity if possible Tapping a tendon to produce a muscle twitch, or stroking the sole of the foot to evoke movement of the great toe are familiar examples Electrical stimuli may be applied directly to peripheral nerves, espe-

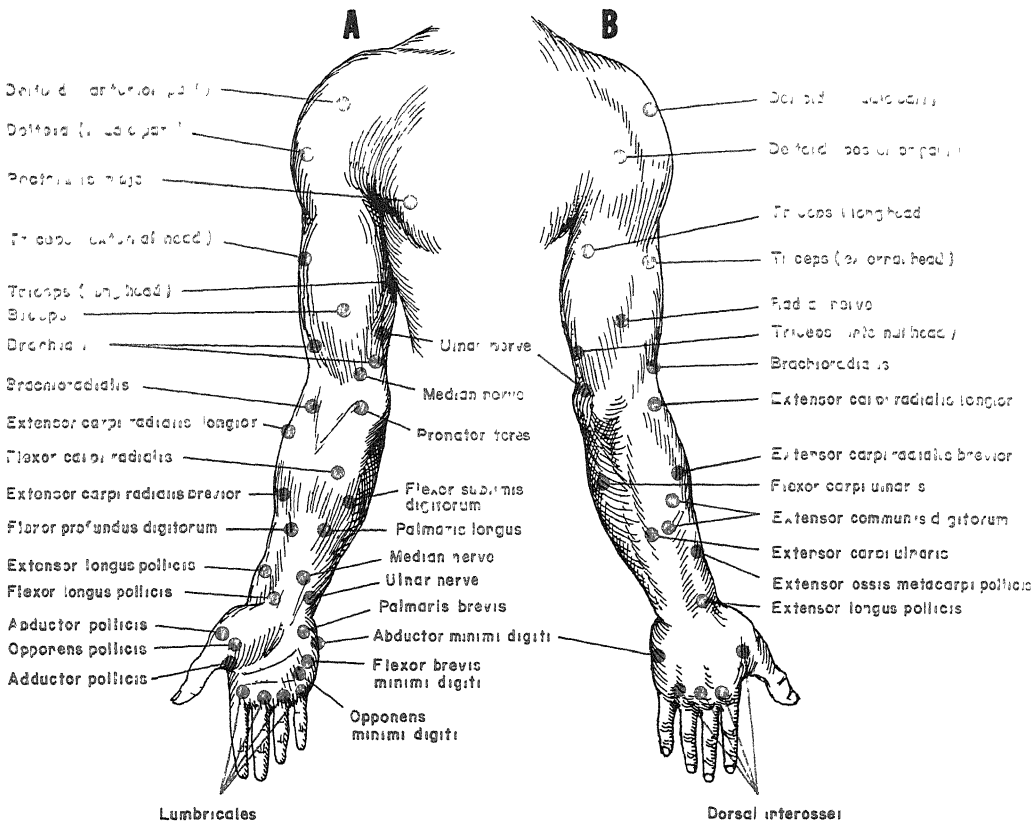


Fig 9 5 Motor points on upper extremity *A* anterior view *B* posterior view Erb cited from Wechsler 1936

cially in areas known as "motor points" where nerves to muscles lie close to the skin (Fig 9 5) Sherrington's experiments using tibialis anticus (previously described) belong to this category of study, the stimulus being delivered to sensory nerves After exposure of the spinal cord by laminectomy, or of the brain by craniotomy, electrodes can be applied to central nervous structures directly This technique is particularly useful for tracing the pathways constituting a system, and thus yields a kind of functional neuroanatomy

Techniques of Stimulation. For several generations of neurologists the stimulating instrument was an induction coil or "inductorium" with a vibrator in the primary circuit to produce alternating current in the secondary Strength was varied by altering the relative positions of the two coils. If a direct rather than an alternating current was desired a battery with a voltage divider was connected through a switch to the tissue, with a device such as a falling weight, a pendulum, or a spring lever arranged so as to open and close switches, and thus to give pulses of known duration Modern stimulators achieve more precise control of frequency, voltage, shape of the voltage or current wave, and duration of each stimulus by use of vacuum tubes or transistors (There is a family resemblance between circuits used in stimulators and those that trigger the sweep of the electronic beam across the screen of a television set) Electrodes, large, small or micro in size, can be applied to the surface or in the depth of the brain by stereotaxic methods On the surface of the spinal cord, on nerve rootlets, or on dissected fragments of rootlets, the electrodes may be applied under either direct or microscopic observation As is the case with a study of injury or ablation, the success of the stimulating technique depends mainly upon skill in examining what is produced by the stimulation Sherrington's results were taken from muscle contractions recorded with levers that measured shortening (for *isotonic* contraction) or tension (for *isometric* contraction) More modern techniques utilize, for example, strain

guages to measure tension, or complex optical systems to measure even the slightest shortening

A relatively new development is stimulation of structures within the brain of human patients in the absence of general anesthesia This was first done by performing the operation under local anesthesia, so that the patient was conscious during the stimulation and could tell the neurosurgeon what he experienced when the stimulus was applied (Penfield and Rasmussen, 1950) Electrodes for this purpose may be manipulated stereotaxically Another but similar technique was developed by Hess (1954) as a device for inserting electrodes through a carrier fastened to the skull (Fig 9 6) Stimulation can be applied without anesthesia, through long, fine wires attached to the electrodes from time to time Or a radio circuit may be employed to transmit without wires to a receiving unit carried by the subject in a collar or harness, with the stimulus applied from a portable unit turned on by the radio signal (Delgado, 1963) The possibility that similar techniques might be used for therapeutic stimulation in human subjects has been considered (Mahl, et al , 1964, Delgado, 1970), with little question that the procedure can be carried out inasmuch as similar devices are now being used with great success as pacemakers for the heart Chronic stimulation of the brain, however, is a somewhat formidable idea because it implies the control of a person's nervous system by someone else This should be noted, therefore, as an example of a scientific procedure that can be made to work technically, but that has not been adopted for general use because of humanitarian or philosophical reservations One should note also that a similar kind of restraint has not been applied to the use of drugs, they are both employed by persons and prescribed by physicians for the sole purpose of altering brain function, often with little attention to possible philosophical objections

Stimulation may be applied by means other than electrical Natural stimulation of sensory receptors may be used, or other types of stimuli may be introduced artificially Thus, thermal, osmotic and chemical stimuli have been introduced into the hypothalamus, brainstem and spinal cord for study of appropriate control systems Drugs may be likewise administered locally by implantation or injection

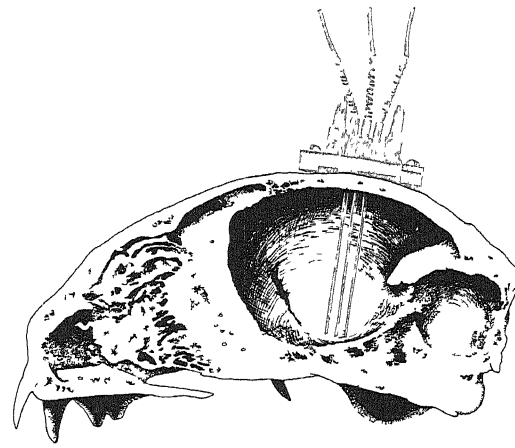


Fig 9 6 Electrode carrier fixed to a sectioned skull to show how insulated electrodes with bared tips can be inserted for stimulation of the brain of a cat without anesthesia The carrier is fixed to the skull during an operation under general anesthesia (Hess 1954)

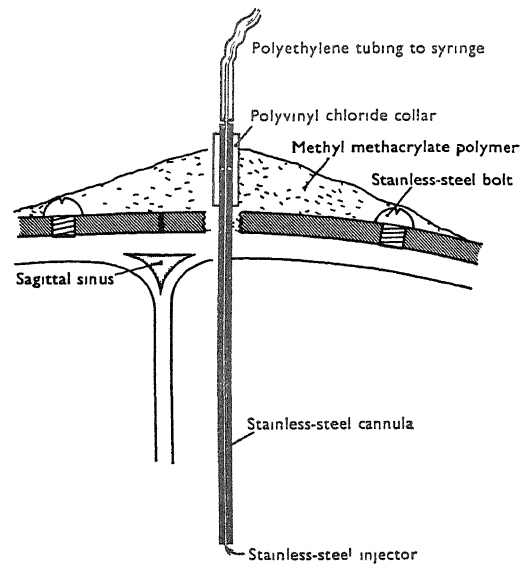


Fig 9 7 Implanted pipette for intracerebral injection of solutions (Epstein et al 1970)

(Fig 9 7) Whereas one drug may act to stimulate a given neural system, another drug may suppress whatever activity may have been present naturally

ELECTRICAL RECORDING

Discovery almost 200 years ago that electricity will stimulate muscle, and that active muscle produces electricity (Galvani, see Mauro, 1969) marked the beginning of parallel

developments in studies of electricity and neurophysiology To detect the presence of an electrical current in muscle Galvani used a nerve to another muscle, activity in the first muscle stimulated the nerve, and hence the second muscle Later authors employed galvanometers of successively improved design, well into this century until Erlanger and Gasser (1937) adapted to their experiments the use of vacuum tubes, amplifiers, and a homemade oscilloscope Modern instruments are solid state descendants of their instruments The important qualities of any recording system are its time constant and its signal to noise ratio (Renshaw et al , 1940) If the amplifier is capacity coupled, with a condenser in series in the input, it will respond quickly to high frequencies but will not faithfully record slow or prolonged wave forms Direct coupled amplifiers, on the other hand, were at first more expensive because of problems of stability and sensitivity, and are somewhat less versatile than the other variety if the strength of the signal is previously unknown Modern instruments, however, are now available commercially with almost any desired characteristics The other property, signal to noise ratio, is important in all electronic instruments. This ratio is significantly greater in amplifiers using transistors than with vacuum tubes, because random electronic processes in tubes produce noise Another consideration in achieving a satisfactory signal to noise ratio is the selection of electrodes that most faithfully gather the potential change in the living system, under conditions where other signals are excluded

Microelectrodes. For early studies of living systems the electrodes were usually metal wires, often platinum or silver, treated so as to minimize polarization of the metal at the point of contact with the tissue Fine tubes or wicks filled with dilute solutions of electrolytes were also used, the electrolytes being those present naturally in extracellular fluid of the animal being studied Woolsey (1958) utilized such electrodes in his studies of the termination of sensory fibers in the cerebral cortex of monkeys When the tubes are fine enough they become micropipettes, and can be utilized for recording potential changes in the neighborhood of single neurons, or even across the cell membrane if a pipette can be inserted inside a cell. In regions where there are generous numbers of large cells, this type of insertion can be accomplished frequently enough to permit studies of changes in membrane potentials associated with stimulation or inhibition of the neuron (Fig 9 8) Many investigators, however, find it more convenient to use sharpened, insulated steel wires as microelectrodes, especially for unit recording from extracellular locations The pipettes used in the experiments of Renshaw et al (1940) had a tip diameter of 15 μ, this measurement is in contrast to sharpened steel needles of some 50 to 100 μ used by Renshaw later (1946), and tip diameters of less than 1 μ for micropipettes of Ling and Gerard (1949) (Fig 9 9)

Electrical Signs of Neural Activity. The fact that electrical changes are associated with neural activity, and the use of electrical or electronic techniques for studying the nervous system, occasionally lead to a confusion that should be avoided The *action potential* is not synonomous with the *nerve impulse*, and the two terms should not be confused As Erlanger and Gasser (1937) phrased it, the action poten-

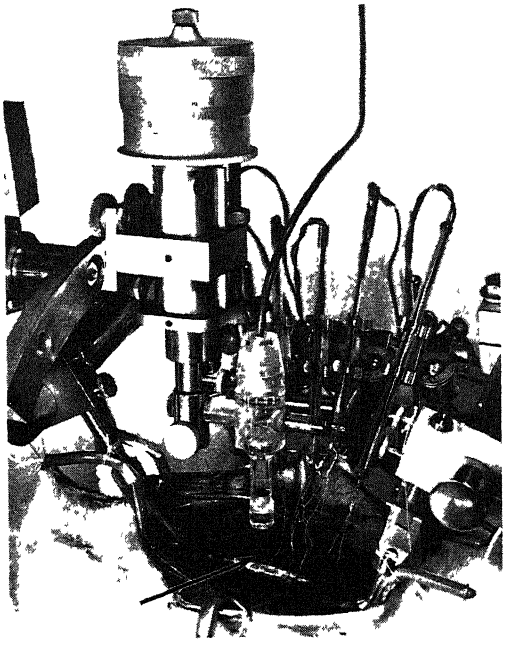


Fig 9 8 Apparatus employed with microelectrodes for intracellular recording from neurons of spinal cord On the left is the rigid frame supportng the micromanipulator the latter carries a preamplifier from which the microelectrode extends (*arrow*) The spinal cord is visible through a pool of paraffin oil that prevents drying of the tissue Other electrodes are for stimulation and other types of recording (Eccles 1957)

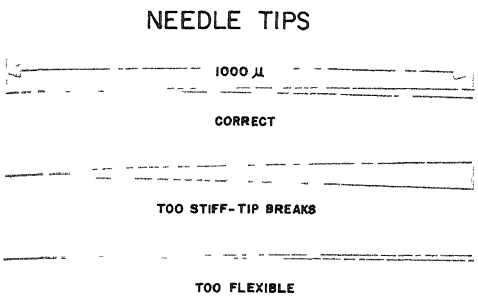


Fig 9 9 Drawing of tips of micropipettes used by Ling and Gerard (1949)

tial is "a sign of nervous activity"—a sequence of potential changes that accompanies the propagation of a nerve impulse What initiates the potential changes is unknown, presumably it is some physicochemical alteration in the membrane of the fiber The fundamental process of the nerve impulse is conduction or propagation of this as yet unidentified change As a result of its propagation along the fiber a wave of sodium penetration is permitted, to be followed in turn by a wave of potassium loss, and then eventually by a wave of restoration of the integrity of the membrane Each of these "waves" is accompanied or signalled by a corresponding potential change and/or current flow

Fundamental processes that enable a nerve fiber to conduct a nerve impulse are presumably the same in peripheral as in central processes of neurons This is in spite of well-recognized differences between fibers within the brain and spinal cord as opposed to those in peripheral nerves (For example, centrally located fibers do not have neurolemma cells, the sheath cells of Schwann The absence of these cells is believed to account for the paucity of effective regeneration of axons in spinal cord and brain) Nevertheless, any recording of action potentials from fibers in spinal cord or brain is believed to include the prepotentials, spikes, and after potentials exhibited by peripheral nerves Yet the shape of the waves recorded from central fibers is different from those recorded in the periphery This difference is caused by peculiaries of the recording site For example, potentials recorded from the surface of posterior white columns of the spinal cord do not have the classic form of an action potential because the transmission of potentials electrotonically throughout the volume of the spinal cord at that location converts what would have been a monophasic record into a triphasic wave (Lorente de Nó, 1947, pp 384–477) Furthermore, the fact that many fibers are active at once, and that they conduct electrotonically the potentials of neurons in their neighborhood or even near their endings makes the potential records all the more difficult to interpret These problems also pertain to potentials recorded from the dorsal or from the ventral roots, where the potentials may reflect conduction in a given fiber, or may arise from changes in neurons lying near or in contact with their terminals, or even potentials in the terminals per se It was to provide some resolution of the uncertainty as to the nature of these dorsal and ventral root potentials that microelectrodes were first employed in the spinal cord by Brock et al (1952) (Fig 9 10) Their use has clarified the nature of synaptic transmission (see sect 1, also pages 9-76–9-82, this section)

Transmission of Control

LOCAL CONTROL PROCESSES

In all types of muscle, and in both exocrine and endocrine glands, the proximate mechanism of control is chemical It is the interaction of some compound or ion with constituents of the membrane of the effector cell that eventually causes contraction or secretion In many, but not all instances, an electrical stimulus can mimic the effect of the nerve impulse

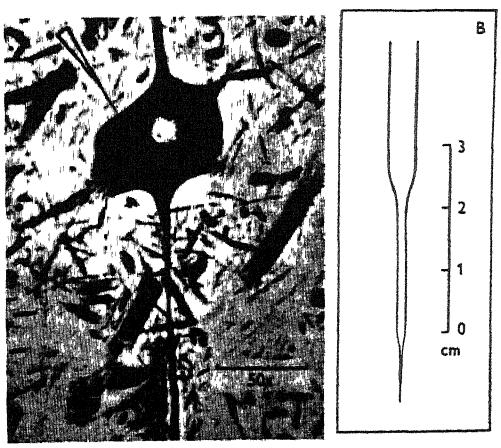


Fig 9 10 Photomicrograph of cat spinal motor neuron and axon (pyridine silver stain), with superimposed tracing of microelectrode tip at same magnification On the right is a scale diagram of the Pyrex microelectrode used in this experiment (Brock et al 1952)

This occurs because electrical currents can bring about physicochemical changes in membranes

If this is true for muscle and gland, is transmission between neurons likewise chemical (or physicochemical)? It is But the truth of this generalization is sometimes obscured by a tendency to describe living systems in physical terms that imply the presence of purely electric or electronic circuits For example, in classical neurophysiology a current is said to "flow" from positive to negative potential sites, whereas in fact any biological potential difference is maintained by differential concentrations of ions, and it is ions moving in or through solutions that carry current The properties of resistance, impedance, conductance, and capacitance as applied to neural systems, and the diagrams that depict membranes as combinations of resistors, condensors, and batteries are only convenient models The resistance of a membrane arises from conditions different from those in a nichrome wire, and the membrane potential is different from the charge stored in a radio condensor Nevertheless, descriptions of the electrical properties of living systems in terms of the properties of macromolecules in watery solutions of electrolytes have been slow to develop What is needed is some way to match the known physical properties of membranes with the molecular and ionic constituents

If proximate control is physicochemical, control at a distance is achieved through an infinite series of similar processes This is, of course, a backhanded way of characterizing a nerve impulse, which is the typical mechanism for neural control at a distance Other types of remote control depend upon the secretion of endocrine glands, the hormones, or other chemicals released into fluids of the body as a result of metabolic processes These, too, induce chemical or physicochemical changes in membranes of cells they control

CODING OF SIGNALS

Approximately 40 years ago Adrian and Bronk (1929) demonstrated that the grading of activity in peripheral nerves depends upon rather simple variables, namely, the frequency of generation of nerve impulses, the interval during which they are generated, and the number of fibers that become active (Adrian, 1935). This means that information is carried from periphery to brain, or from brain to effectors, in the same way as in a digital computer A nerve impulse is one "bit" of information, and is either propagated or not propagated The information carried by a nerve, then, is first of all a function of the number of bits or nerve impulses (Figs 9 11 and 9 12) Another possibility of information transfer is in the patterning of nerve impulses (Bullock, 1967) When a message is sent by telegraphy, for example, the operator does not simply count the number of pulses, he unconsciously analyzes their pattern From dots, dashes, and

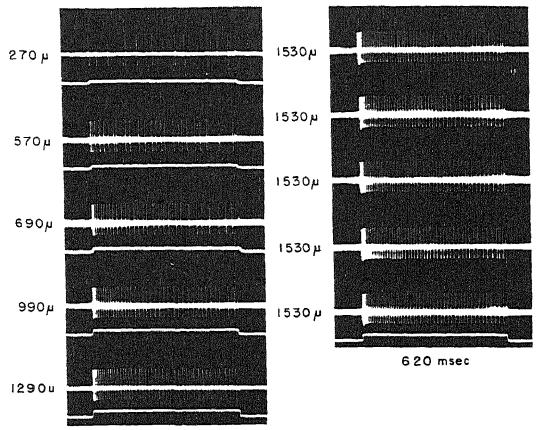


Fig 9 11 Frequency of discharge of single sensory neuron in response to indentation of skin Upper record of each pair is action potentials lower record shows indentation stimulus Depth of indentation is given in micrometers on ordinate From median nerve of a monkey (Mountcastle 1967)

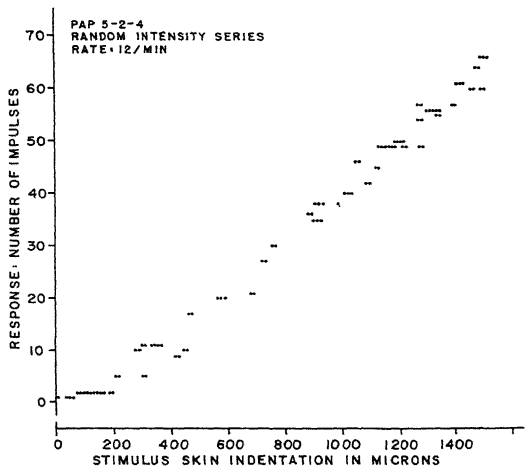


Fig 9 12 Plot of data from experiments such as those of Fig 9 11 Ordinate is derived from the record of nerve impulses Cited from Mountcastle 1967

pauses he recognizes characters of the Morse code A nerve cannot carry the equivalent of a dash, but it can transmit patterns of bits that represent some biological coding It may be that this coding involves a frequency modulation of the impulse pattern, or some other mechanism still more subtle and elegant in its operation (Mountcastle, 1967)

Although transmission of information in the nervous system is digital, mechanisms of computation in spinal cord or brain are more like those of an analogue computer When nerve impulses arrive at a central terminal they bring about quantal, minute release of a transmitter that then creates or contributes to a state that is not unitary, but that can be infinitely graded upon the membrane of the postsynaptic neuron (Fig 9 13) (A similar process occurs at receptors before and while they are stimulated) Whether this change is electrical or physicochemical is not of greatest importance because of the dependence of one on the other, as previously noted What is significant is that the integrative actions of the nervous system are based upon transmission of bits of information, the conversion of bits into time-dependent changes of varied intensity, and finally the influence of this change upon cells or terminals in the region where it occurs Ordinarily this active region is a small locus where the terminal of a nerve fiber lies practically in contact with the membrane of the next element in the pathway

CENTRAL COMPUTATION

The nature of the computational processes at central terminals is being actively studied in many laboratories What may be involved was illustrated long ago in the analyses Sherrington and his associates made of spinal reflexes (Creed, et al , 1932) When they demonstrated that even at the spinal or segmental level the spinal cord contains mechanisms capable of both exciting and inhibiting succeeding neurons in a reflex arc, they gave names to the reactions set up within the spinal cord by the afferent impulses For this purpose they used the analogy of what happens in a nerve when it is stimulated electrotonically, that is, by passage of a direct current Such a current creates a *catelectrotonus* under the negative terminal and *anelectrotonus* under the positive The catelectrotonus is characterized by what has been called a local excitatory state (LES) that is defined as a tendency of the nerve to initiate a nerve impulse in that location The LES has the property of being conducted along the fiber by electrotonic mechanisms rather than by the conducting system responsible for the nerve impulse Consequently, the LES falls off sharply in intensity with distance from the cathode Furthermore, the LES is not an all-or-none phenomenon, in that it can be made stronger by intensifying the current, or weaker when current is reduced Inasmuch as the anelectrotonus under the anode has the opposite kind of properties, one sees that the LES may be said to have either a positive or a negative sign, tending either to excite or to depress excitation of the fiber Finally, when the current is turned off the LES decays exponentially, and is not followed by the refractory state that follows conduction of a nerve impulse



Fig 9 13 Electrical control of frequency of miniature end plate potentials (upper record of each pair) during passage of a current with cathode near the end plate (lower records) Each miniature potential is associated with a quantum release of transmitter although resulting overall effect on postsynaptic membrane is graded in essentially infinite number of steps (del Castillo and Katz cited from Katz 1971)

Using the local excitatory state as a model, Sherrington and his colleagues supposed that an analogous situation is created in the spinal cord by the arrival of impulses via afferent nerve fibers (Fig 9 14) For the changes involved he utilized the terms, central excitatory state (CES) and central inhibitory state (CIS) so as to identify the influence upon the postsynaptic neurons The analogy between LES in the periphery and what happens centrally is clear enough, but there is one significant difference In spinal cord and brain the change having the negative sign, that is, the inhibitory process or CIS, is much more significant than any similar change heretofore discovered in peripheral nerve or receptors of higher animals and man This point cannot be overemphasized

In spite of intensive study, the nature of the

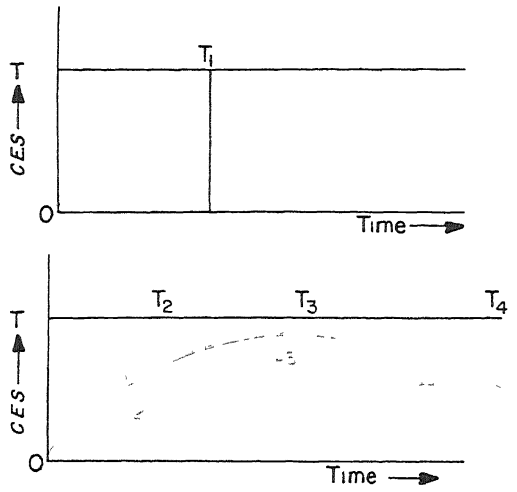


Fig 9 14 Hypothetical time course of central excitatory state (CES) *Upper record* T-T' interval is latent period *Lower record* When a second stimulus is applied at point S_2 S_3 or S_4 after an initial stimulus not strong enough to induce a nerve impulse the latency of the response to the second stimulus depends upon the magnitude of the CES at that time (Creed et al 1932)

CES and CIS remains unknown It is known that they have electrotonic components, as reflected in potentials recorded from either dorsal or ventral roots Barron and Matthews (1938) proposed the hypothesis that the CES is, in fact, a central electrotonic state like peripheral catelectrotonus (Fig 9 15) In their view the CIS is the removal of previously existing CES Eccles (1957) and others have confirmed their observations, and using microelectrodes have recorded changes in transmembrane potentials during excitation and inhibition Nevertheless, there remains much uncertainty as to just how the arrival of nerve impulses in afferent fibers can lead to these alterations in membrane potentials and thus give rise to the potentials that are conducted electrotonically out through the spinal roots

The most popular hypothesis to explain CES, CIS, and the characteristics of synaptic transmission is that the processes are chemical in nature, brought about by discharge of a compound that is released from nerve endings, crosses the minute space to the postsynaptic membrane, there to act upon specialized receptor sites so as to induce changes in membrane potential and in excitability (sect 1, chap 3) Evidence for this opinion is impressive, although indirect In the first instance it is taken by analogy from studies of transmission in peripheral parts of the autonomic nervous system, where acetylcholine, epinephrine, and norepinephrine are known to be responsible for the transmission process The fact that the central nervous system contains these compounds, especially acetylcholine and norepinephrine, and also serotonin (5-OH tryptamine), dopamine (dihydroxyphenylethyl amine) and other compounds known to be active in peripheral systems, plus γ aminobutyric acid (GABA) as a candidate for the role of central inhibitor, is taken as support for the concept that transmission occurs by similar mechanisms both peripherally and centrally Furthermore, the localization of these compounds, where it is known, is in the endings where transmission is initiated (de Robertis, 1964, Peters et al , 1970) Finally, neurons in spinal cord and brainstem can be excited or inhibited by at least certain of these compounds injected into their neighborhood, and typical changes in membrane potential have been reported under these conditions (Grundfest, 1967)

The possibility that the central computational processes are chemical in nature with a variety of substances serving as transmitters is an intriguing one, because it multiplies the number of conceivable types of central control

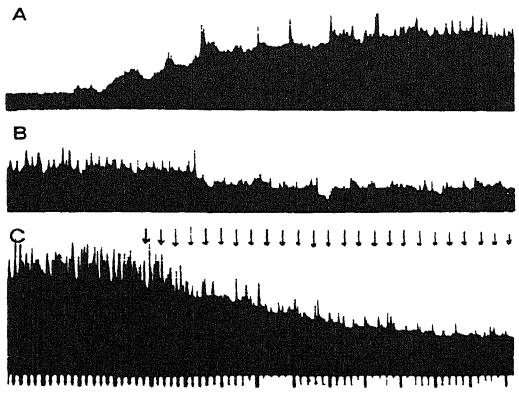


Fig 9 15 Electrotonic and action potentials recorded from anterior rootlet of cat (*A* and *B*) and frog (*C*) Heavy black area at bottom of each record is measure of electrotonic potential on which action potentials are superimposed *A*, imposition of reflex excitatory stimulus *B* addition of reflex inhibitory stimulus *C* Potentials evoked by a weight on foot inhibited by contralateral reflex stimulation at each arrow (Barron and Matthews 1938)

This concept leads to the possibility of selective CES and CIS, in addition to actions that are simply plus or minus With perhaps half a dozen transmitters acting at millions of loci, all having inputs from a variety of downstream loci, the resulting permutations and combinations of circuits offer limitless possibilities for controls and integrative actions

Hierarchy of Control Functions

SIMPLE CONTROLS

There are a few instances where control mechanisms are localized within tissues or organs, and do not require intervention from the nervous or endocrine systems Perhaps the best known example of this in higher animals is the action of acid metabolic products including CO_2 in increasing blood flow through small vessels of active muscle The next higher kind of control is found in invertebrates (see Wiersma, 1967), where reflexes are completed in ganglia that are not a part of the "brain" of the animal In higher animals including mammals there is no clear evidence that this type of ganglionic reflex is possible, although many investigators have tried to establish its presence in autonomic ganglia such as the coeliac and superior mesenteric These efforts have been inconclusive, and the best evidence is that information from detector mechanisms can influence ganglion cells only after transmission to the spinal cord and back to the ganglion via efferent nerves

In a moment the conduct of information to the brain via non-neural routes will be described So far as neural mechanisms are concerned, however, all information brought to spinal cord and brain via nerves arrives through only a few pathways—the dorsal roots of spinal nerves, the fiber tracts functioning as cranial nerves I (olfactory) and II (optic), and afferent fibers in cranial nerves V, VII, IX, and X Similarly, all neural control leaves the brain and spinal cord via either the spinal nerves or the cranial nerves (with the possible exception of nerves I and II) This pattern of neural input and output is common to all vertebrates, what differs among the several phyla, orders and species is the complexity of interconnections between input and output, and hence the complexity of the integrations they are able to sustain.

LEVELS OF CONTROL

From the preceding generalization it follows that one should be able to predict the complexity and variety of possible activities from a knowledge of the number of interneurons and their interconnections (Bullock and Horridge, 1965, p 312) Such a prediction is useful, not only in comparative neurophysiology, but also in comparison of functions of different portions of a single nervous system For example, the spinal cord of lumbar levels seems relatively simple in organization (p 9-21), with segmental mechanisms subserving reflex responses, and with appropriate connections to higher and lower mechanisms But the lumbar cord as confined in the semiflexible vertebral column could not possibly contain the control systems required for a man's walking or running The additional neurons required for such activities are located in the larger parts of the brain, the cerebellum, basal ganglia, and cerebral cortex From this, one can infer the hypothesis that local controls tend to be of the simplest type, with complexity added from higher levels—inasmuch as it is the higher levels that have the extra neurons This principle was explicitly stated by J Hughlings Jackson (1890) in a lecture summarizing his studies of convulsive seizures He identified three levels of control, as follows

"(1) The lowest or first level is roughly and incompletely defined as consisting of cord, medulla and pons, and more completely, and yet still roughly, as being that sensorimotor division of the central nervous system, to and from which pass nerves for every part of the body

"(2) The middle or second level (its motor province) is composed of centres of the Rolandic region (so-called 'motor region' of the cerebral cortex), and, possibly, of the ganglia of the corpus striatum also

"(3) The highest or third level is made up of centres of prefrontal lobes [and] represents most complex movements of all parts of the body from eyes to perineum "

In considering the nervous system 80 years later, one appreciates the clarity with which Jackson perceived this organization, but is inclined to modify it by suggesting the presence of five, rather than three levels, as follows

1) Segmental level, where reflex controls of spinal cord and brainstem do not depend upon

neurons located outside the main sensory, internuncial, and motor columns,

2) Reticular level (named from the reticular formation of the brainstem, see p 9 31), where selective facilitation and inhibition modify both phasically and tonically the activity of segmental reflexes,

3) Level of unconscious control of exchanges, including pulmonary ventilation, water balance, body temperature and energy exchange, and involving mechanisms of brainstem and diencephalon,

4) Level of voluntary control of movement of skeletal muscles and of projection of principal sensory systems—cerebral cortex, basal ganglia, thalamus, and cortico-pontile-cerebellar pathways that give voluntary motion its customary and normal character,

5) Highest level, presumably cortical, where neural activity is not necessarily translated into motion or secretion

The first four of these levels have a common property, each succeeding higher level utilizes mechanisms resident in lower levels Whether this is true of the highest level is not known As many authors have pointed out, one can analyze neural functions only in terms of some response—some output If a response is excluded by definition of the experiment an observer cannot know whether or not neural activity is taking place We know that someone is "thinking" only when he tells us so

NON-NEURAL INPUT SIGNAL GENERATION

Outside, and yet in parallel with, these levels of neural function is another means of communication provided by the blood supply to brain and spinal cord Its significance is sometimes overlooked It is the means by which neural activity can be modified, and probably even initiated, as a result of relatively nonspecific changes extraneous to the pathways of nerve fibers That this can occur has been shown most clearly in the cases of regulation of body temperature, water balance, and pulmonary ventilation Warming the blood, or changing its osmolarity, or altering the pressure of CO_2 seems to activate mechanisms lying in the brain (If the phenomenon is not truly activation, it is at least major modification of neuronal output) Perhaps this only shows that the brain contains specialized detector mechanisms similar to those located in the periphery—for temperature and for taste, for example Nevertheless, the disclosure that neuronal output can be initiated and/or modified by this type of humoral changes raises the possibility that other kinds of changes brought by the blood may have similar actions on other portions of the brain There is evidence that portions of the hypothalamus are sensitive to circulating hormone levels (sect 7, chapts 2 and 9) Perhaps other neurons including those of the cortex have similar sensitivity This may be the manner in which behavior associated with reproduction, for example, is initiated by the central action of sex hormones (Young, et al , 1964)

The output of a neuron that responds to some nonspecific variable such as P_{CO_2}, osmolarity or temperature, is analogous to the performance of an electronic circuit known as a *signal generator* (Strumwasser, 1967) Such a generator may convert a steady, DC voltage into pulses of electrons Circuits to do this are commonplace in electronic devices, including musical electronic organs where the notes may be produced in this fashion by oscillators As defined in this way any detector that responds to a stimulus such as pressure or stretch is a signal generator The output of such a cell is a rhythmic train of nerve impulses Although it is commonplace in sense organs, this type of neuronal activity is not widely recognized in discussions of the central nervous system, where most authors have a tendency to emphasize the stimulus-response relationship dependent upon neural as opposed to humoral or generalized input Perhaps this explains the seemingly endless character of the debate over the question of whether every neural output is a response to some identifiable input That question is most important in its relevance to "voluntary" or "spontaneous" motor activity Can a movement be truly voluntary, or is it necessarily a response to some previous stimulation? If it is true that the central nervous system does contain signal generators, then the question of spontaneous activity takes on a new significance

However this may be, microelectrode recordings from brainstem and hypothalamus reveal what appears to be central detection, and hence signal generation This phenomenon has been utilized to explain the respiratory "drive," and it seems probable that it can be extended also to the waking state, to consciousness, and to forms of behavior that include locomotion, feeding and drinking

Significance of Glia. Intermingled with

neurons in the brain and spinal cord there are other types of cells classified together as glia One species, the *neuroglia*, originates from ectoderm as do the neurons, it is found in two subclasses, the *astrocytes* and the *oligodendroglia* The other, known as the *microglia*, is of mesodermal origin and enters the embryonic brain when it is penetrated by blood vessels (Fig 9 16)

A variety of functions has been attributed to glia, including the supporting of neurons, their nourishment in some trophic manner, and the repair of injury They must contribute to the biochemical and biophysical milieu in which the neurons live (de Robertis and Carrea, 1965, Bodian, 1967, Trachtenberg and Pollen, 1970) In a sense the neurons are subjected on the one hand to shifting concentrations of materials in the plasma and interstitial fluid phases within the brain, and on the other to relatively stable rates of exchanges carried on by glia One interesting possibility is that the constancy of the glia might function as a setpoint for regulations carried on by neurons For example, if the rate of metabolism, and hence of heat production of glia are constant, this might provide a setpoint for regulation of body temperature Theoretically, this idea might be extended also to other chemical and physical properties of the environment of neurons as determined by the glia in their neighborhood

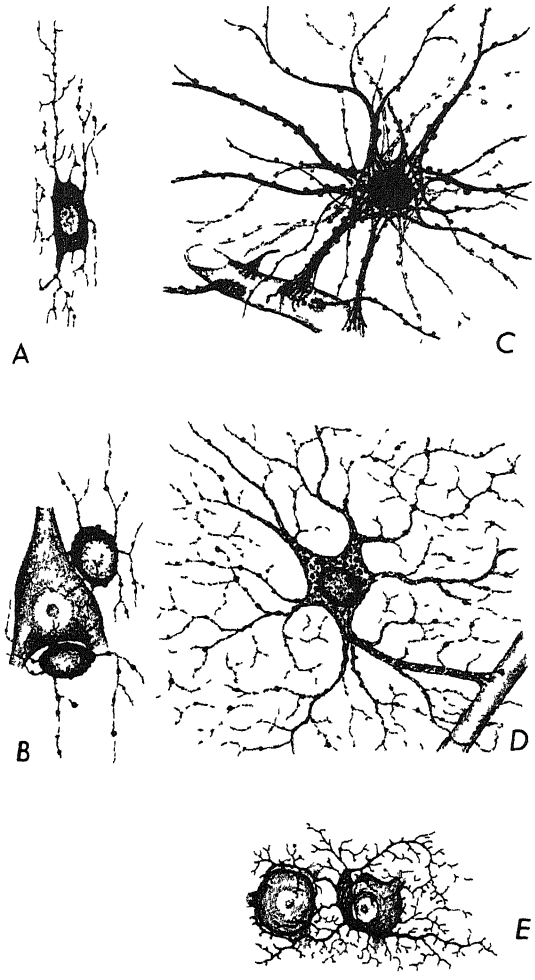


Fig 9 16 Drawings of typical glial cells *A* oligodendrocyte *B* two oligodendroglia lying against a neuron *C* fibrous astrocyte with processes on blood vessel *D* protoplasmic astrocyte *E* microglial cell in vicinity of two neurons From Penfield as reproduced by Copenhaver et al 1971

PRIORITY AND COMPROMISE

If there are many levels of controls and only a few motor outputs, one must ask just how does the efferent component know what it is supposed to do? What determines whether a given set of neurons is employed for posture, for walking, for getting food, for copulation, for building a fire, or for composing textbooks? If the nervous system is considered to be like a telephone system, with the telephone instruments representing sites of both input and output and with exchanges serving connector and computational functions, there at once appears a notable difference between the telephone and the nervous systems When a telephone line is in use, it cannot be secured for another purpose no matter how urgent One can imagine, however, a more elaborate version of a telephone system where a given line becomes available to whatever caller has the greatest need for it, or thinks he has In small towns of 50 years ago the telephone system worked in this fashion, because one could try to persuade the "operator" to interrupt another conversation so as to make a line available for some critical call This is what the nervous system always does, it accommodates itself so as to maintain certain priorities of action, and in doing so often makes compromises Pulmonary ventilation is compromised during speech, especially during singing; energy balance is not maintained during mating in many species, nor during exposure to high environmental temperature in homeotherms But more commonly the compromise is, for example, that mechanisms preserving posture are altered during movement, that activity in an efferent system is suppressed for movement

of its antagonist, that what might become a prolonged movement is converted into a phasic or rhythmic alteration In this sense there is a hierarchy among even spinal reflexes, with some serving higher integrations than others (Sherrington, 1906) Nevertheless, even the compromises are brought about, and priorities are established, by means of the CES and CIS evoked in central mechanisms as a result of incoming neural activity, or in response to information transmitted via the blood stream It is the characteristic localization, distribution, and intensity of the CES and CIS that is essential for every variety of central neural activity, even the most elaborate, intricate, or subtle

Control vs Regulation

DEFINITIONS AND EXAMPLES

Thus far in this chapter the word *control* has been used in its common English meaning This same meaning, however, can have a technical significance if control is distinguished from a related word, *regulation* In physiology the latter implies more than the former, because it is used by many authors to signify that controls tend to preserve some almost constant condition For example, the action of the heart is controlled, in part by nerve fibers from the vagus and from sympathetic nerves But blood pressure in the aorta is regulated, because these and other control systems keep it at an almost constant level There are a very few instances in physiology where a constancy may occur solely as a result of interaction of physical forces This happens, for instance, when a resting, fasted man is exposed without clothing to an environmental temperature of about 28° C Under these stated conditions his "basal" rate of heat production will equal his rate of heat loss to the 28° C environment Usually, however, a constancy of a physiological variable requires a feedback mechanism that takes information about the output back to the control system so as to modify the nature of the control Feedback loops are well known in mechanical and electrical devices The governor on a steam engine, a thermostat controlling a furnace, as well as the control of an air conditioner all depend upon feedback loops In driving a motor car at a constant speed, watching the speedometer is the feedback component of the control system Analogous mechanisms generate regulations in living organisms Lavoisier (see Adolph, 1961) and Claude Bernard (1878) long ago recognized that life depends upon these regulations, and that they function to maintain a relative constancy of composition of the fluids of the body They also maintain other constancies, including body temperature, and have profound effects upon behavior through their ability to direct and even to dominate the activity of all parts of the neural and endocrine control systems

According to this concept, then a *regulation* utilizes control systems to achieve a relative constancy as a result of *detection* of some variable and the interposition of a *feedback loop* between output and input (Fig 9 17)

Because skeletal muscle contains detectors that initiate a feedback of information to the spinal cord, one may ask whether the action of skeletal muscles is regulated or controlled It is a regulation that prevents slackness of muscles, either at rest or during movement Through feedback from muscle spindles (p 9-82), and with the collaboration of γ efferent fibers back to the spindles, each muscle is maintained at a length appropriate to the position of the joint it moves With reference to the control of movement, however, the concept of regulation becomes less applicable, in part because of problems that arise in the analysis of a dynamic (as distinct from a static) condition One can identify the possibility of regulation of motion in a soaring bird, perhaps, or in a fish maintaining its place in a moving stream, or in the unconscious standing posture of almost any subject The rigid caricature of standing seen in a decerebrate animal in a laboratory (p 9-101) is often given as an example

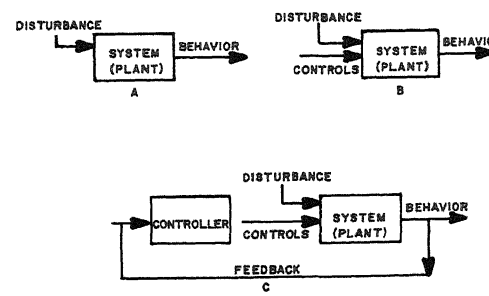


Fig 9 17 Development of feedback control patterns *A*, uncontrolled system *B*, open-loop control *C* closed-loop control with feedback Grodins, in Yamamoto et al 1969

of a postural regulation, maintained by feedback mechanisms from stretch receptors, joint receptors, and the labyrinths, acting upon control systems in the brainstem Nevertheless, the concept of regulations is not yet ready to try to account for the fleeting, shifting, fluid movements of a cat's normal gait

There is, however, another very interesting relationship between the concept of regulation and voluntary movements These movements are utilized in regulation of body temperature, drinking, food intake, and even in pulmonary ventilation It is possible voluntarily to alter a given regulation for a brief period of time, although at the risk of inducing air "hunger," thirst, or sensations of chilling or overheating Ordinarily the regulation will be reinstated and the voluntary disturbance will be counteracted Conversely, regulations often alter voluntary activity—in fact, the dominance of the body's regulations over many forms of voluntary activity seems to be a condition required for preservation of life of an individual as well as of a species

SETPOINT

When controls produce a regulation with some constant value, one wonders how the system "knows" what this value should be Moreover, in many biological systems the value changes from time to time, for example, normal body temperature is lower than the temperature of a working individual, while in fever a completely different level may be maintained (Hammel, 1965) In engineering systems a setpoint requires the generation of some type of constant signal against which the feedback signal may be compared Whether the nervous system is capable of generating such a reference signal is a problem of considerable interest (Hardy, 1965) In biological systems there is another possibility, however, that has a significance different from the conventional idea of a setpoint Consider the circulation and regulation of blood pressure It is conceivable that the upper limit of aortic blood pressure is determined merely by the sensitivity of stretch receptors in the carotid sinus and arch of the aorta If the pressure goes high enough to stimulate the receptors, reflexes will be initiated to lower the pressure The higher the pressure the more active the receptors, and thus the more powerful the reflex responses to bring the pressure down Similarly, if there is somewhere in the body a detector that responds to a reduction in blood pressure, with a threshold of hypotension at which it becomes active, reflex controls to raise pressure can be set into action under this set of conditions In this model the nervous system that does the regulating is not required to contain a generator for a setpoint, the constancy of pressure being a result of the alternate action of hypertensive and hypotensive detectors and the controls they bring into play Thus, conceptually the inherent sensitivity of the detectors replaces the setpoint generator It follows that according to this model the level at which the constancy is preserved is a function of the sensitivity of the detectors Survival of a species shows that the essential detectors for all of the vital regulating systems have sensitivities in the appropriate range

A final word may be added about how detectors acquire and maintain this particular sensitivity In mammals it seems likely that their sensitivity is determined by the composition of the environment in which they develop The detectors of the fetus are exposed to compositions determined by the mother's detectors There is reason to believe that detectors have an intrinsic property of showing least activity when the variable to which they respond has a value that is unchanging This phenomenon is known as *adaptation* in nerve endings (sect 8, chap 1) or as *accommodation* in nerve fibers (sect 1, chap 3) It may be that chemoreceptors in the fetal arterial system, for example, acquire a sensitivity that reflects in some measure the CO_2 and O_2 pressures in the system in utero at the stage when these control reflexes became specialized Alternatively, their sensitivity might be determined in the neonatal period as a kind of physiological imprinting In any event, there presumably is a continuity of controls and regulations from one generation to another on a genetic basis that in itself might be the means of determining sensitivity of the critical detectors

2 ORGANIZATION OF THE CENTRAL NERVOUS SYSTEM

General Considerations

All parts of the nervous system of mammals including man originate either from the neural tube or from the neural crest of the embryonic ectoderm The *neural crest* gives rise to the neurons whose cell bodies are located in dorsal root ganglia and homologous ganglia on cranial nerves, it also produces the postganglionic neurons of the autonomic motor system (Fig 9 18) The *neural tube* acquires landmarks that permit the spinal and cranial portions to be distinguished as early as the 3rd week in a human embryo Nevertheless, even in the adult there is no significant difference in structure between the medulla oblongata just above and the spinal "medulla" or cord just below the foramen magnum Distinction between the two is based on anatomic relations rather than on any unique feature of their organization

EMBRYOLOGY OF THE BRAIN

The cranial part of the neural tube becomes further demarcated into a series of five "vesicles," or fluid-containing swellings, that eventually produce the neurons of every part of the brain How the brain develops from these vesicles, where neurons of the several parts come from, and especially how it flexes upon itself (Fig 9 19) and comes to occupy the space available in both anterior and posterior fossae are fascinating topics, albeit too detailed for consideration here Briefly, the cerebral cortex becomes more and more voluminous as it differentiates with its underlying fibers and basal ganglia, until the cortex comes to envelope the brainstem as rising bread dough may completely overflow and eventually hide the pan in which it was mixed To a lesser degree a similar expansion occurs in the cerebellar vermis and cortex It is this overgrowth of the cerebral cortex in every direction that explains certain puzzling relations (Fig 9 20) For example, if a cadaver brain is sliced in the coronal plane, the section may disclose a number of apparent cavities or spaces (Fig 9 33, below) These may include the body of the lateral ventricle, its inferior horn in the temporal lobe, the third ventricle, and perhaps the foramen of Monro connecting it on either side to the lateral ventricle The puzzle, however, is the presence of a cleft that lies between the lateral and third ventricles, but that is separate from both of them It is the transverse fissure of the cerebrum, continuous with the subarachnoid "space" that everywhere surrounds the surface of the brain, and a token here of how the diencephalon has been "buried" by overgrowth of the telencephalon

NOMENCLATURE

Names used to designate the several major portions of the central nervous system usually describe their appearance or their anatomical relations Fiber tracts are known eponymically after the anatomists who discovered them, or more precisely are given the names of the parts they join—the corticospinal pathway, for example Another type of designation comes from the five embryonic brain vesicles, and lends itself to the concept of levels of function as presented in the preceding chapter In this terminology the origin and level of each of the major components may be summarized as follows

Telencephalon (end brain)
- Cerebral cortex
- Central white matter of the cerebrum, corpus callosum, internal capsule
- Basal ganglia claustrum, caudate nucleus, globus pallidus, putamen, amygdala
- Hippocampus

Diencephalon ("in between" brain)
- Thalamus
- Hypothalamus
- Epithalamus pineal gland, stria medullaris, habenular region and related structures

Metathalamus medial and lateral geniculate bodies
Subthalamus (ventral thalamus)
Mesencephalon (middle brain)
Corpora quadrigemina (superior and inferior colliculi)
Cerebral peduncles, substantia nigra
Tegmentum of mesencephalon
Metencephalon ("after" brain)
Cerebellum—cortex and nuclei
Tegmentum of pons
Pons
Myelencephalon (marrow brain)
Medulla oblongata

Much of the material that students are expected to learn in neurophysiology is, in fact, applied neuroanatomy Courses in the latter subject usually include information about function, because structure and function are necessarily linked, and learning one aids in remembering the other For this same reason an account of how the nervous system works and what it does cannot omit certain details of how it is put together Consequently, a brief review of the major components of the central nervous system and their connections is given in this chapter Students who have previously studied neuroanatomy can well afford to go directly to Chapter 3 For additional information about the anatomy of the system, Truex and Carpenter (1969) may be utilized, or another similar textbook Clarke and O'Malley (1968) have published a most valuable collection of citations to classic work in neurophysiology and anatomy

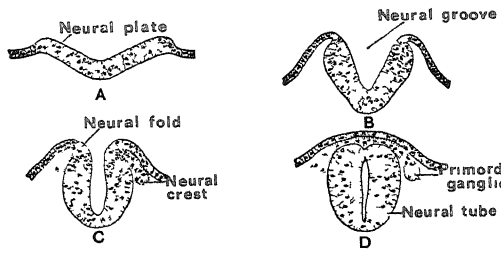


Fig 9 18 Development of neural tube and ganglia from ectoderm of neural plate and neural crest in human embryo (Arey 1965)

STRUCTURE AND RELATIONS OF THE SPINAL CORD

The *gray matter*, centrally placed, is composed of a mass of nerve cell bodies and nerve fibers (dendrons and axons), mostly unmyelinated, supported by a framework of neuroglia It is well supplied with blood vessels In the ventral columns are situated the large, multipolar bodies (100 μm in diameter) of the α motor neurons whose axons leave the cord by the ventral roots (Fig 9 21) Each axon ends on a group of skeletal muscle fibers—the neuromuscular structure constituting the so-called motor unit The small or γ cells of the ventral columns innervate the intrafusal fibers of muscle spindles, and thus control the threshold of these important proprioceptive end-organs In the thoracic and upper lumbar segments, the gray substance lying between the ventral and dorsal columns shows a small lateral projection This is the lateral column or horn, it contains a cluster of nerve cells (the *intermediolateral cell column*) that give origin

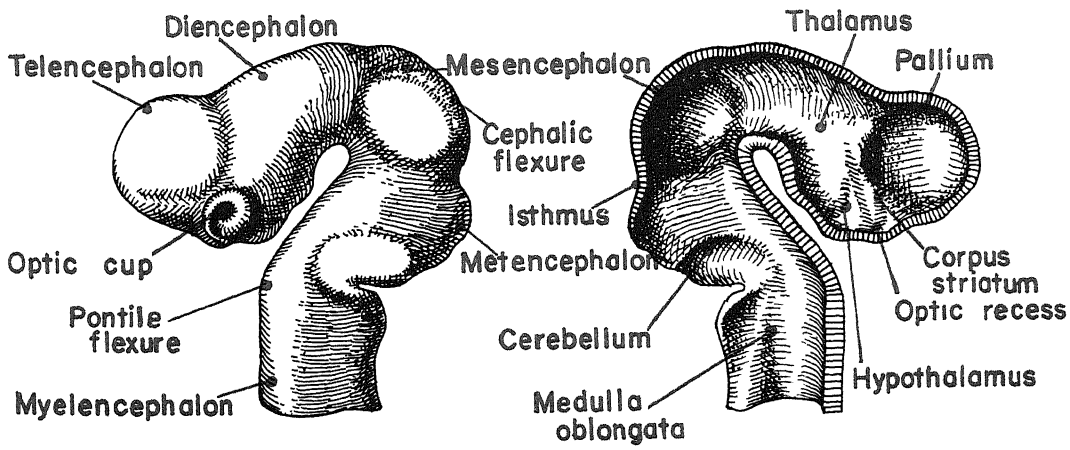


Fig 9 19 External and internal views of vesicles of embryonic human brain His Prentiss-Arey from Ranson and Clark 1959

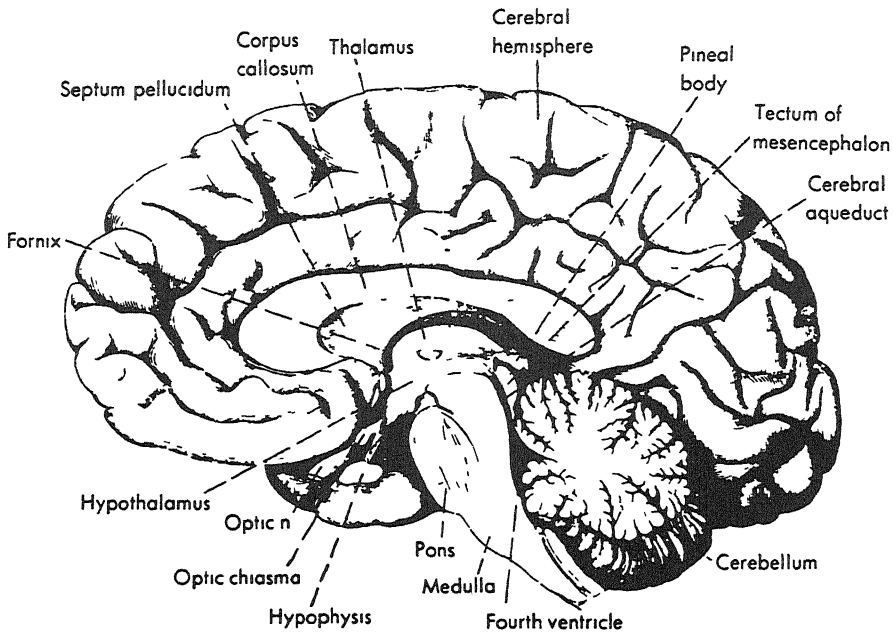


Fig 9 20 Medial sagittal section of the human brain From Henle 1868 reproduced from Hollinshead 1967

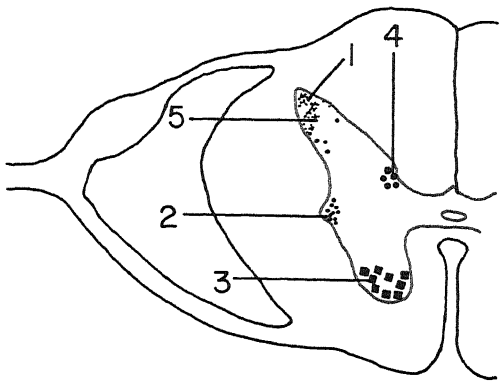


Fig 9 21 Location of principal cell groups in grey substance of thoracic spinal cord

to the preganglionic, autonomic fibers which also leave the cord via the anterior (ventral) nerve roots Functionally these cells are connector or internuncial neurons The well-defined collection of cells occupying the inner part of the base of the posterior horn is known as *Clarke's column*, this group, since it is confined almost entirely to the thoracic region of the cord is also known as the *dorsal nucleus* It is homologous with the nucleus cuneatus of the medulla oblongata To the outer side of the base of the posterior column is an area where strands of white matter and prolongations from the main mass or gray matter intermingle to form a delicate interlacement This is known as the *reticular formation* (formatio reticularis) and is most prominent in the cervical region It is continuous with the reticular formation of the medulla oblongata and pons At the apex of the dorsal horn is a cap of what appears to be gelatinous material containing groups of small nerve cells possessing many dendrites and named the *substantia gelatinosa of Rolando* It is believed to receive the finest fibers (myelinated and unmyelinated) of the posterior roots, and rostrally is continuous with a sensory column of the fifth nerve, the "spinal nucleus" The remaining part of the substance of the dorsal horns constitutes the *chief sensory nucleus* The bar or isthmus connecting the two lateral masses of gray matter across the midline contains the *central canal*, derived from the enclosed cavity of the embryonic neural tube, it is continuous with the ventricular system of the brain The parts of the gray isthmus in front of and behind the central canal are referred to as the *anterior* and *posterior gray commissures*, respectively

The surrounding *white matter* is composed of bundles of fibers both myelinated and unmyelinated, the former predominating Each half of the spinal cord is marked out by the fibers of the ventral and dorsal nerve roots into three white columns or *funiculi—ventral* (an-

terior), *lateral* and *dorsal* (posterior) The band of white matter lying in front of the gray commissure is called the *anterior* (ventral) *white commissure*

The Spinal Nerve Roots

There are 31 pairs of spinal nerve roots, a pair from each spinal segment The anterior and posterior roots join within the vertebral canal just beyond the posterior root ganglion to form a corresponding number of spinal nerves

The *anterior roots* of the spinal nerves are composed entirely of efferent fibers These are the α and γ myelinated fibers to skeletal muscle, and also, in the thoracic and lumbar regions, the fine, lightly myelinated preganglionic fibers of the autonomic nervous system The *posterior roots* are constituted of afferent fibers from the skin, muscles, and viscera They are the central processes of the unipolar cells of the spinal ganglia There is no important evidence that they contain efferent fibers The generalization that the dorsal roots are afferent and the ventral roots efferent in function is called the *Bell-Magendie* law Two divisions of the posterior root, a medial and a lateral, are distinguished The lateral division is largely composed of small unmyelinated fibers, these, after entering the cord form the small triangular area at the tip of the posterior horn known as Lissauer's tract The fibers of the medial division enter the posterior columns of the cord (see Fig 9 22)

SEGMENTAL DISTRIBUTION OF THE SPINAL NERVES

In the young mammalian embryo and in certain adult lower forms, e g , fishes, the body is demarcated into a regular series of transverse segments or *metameres* The muscles (*myotomes*), skin (*dermatomes*) and viscera of each of these primitive blocks receive innervation from the nerve roots of a corresponding spinal segment The anterior root of each spinal nerve supplies motor fibers to the respective myotome, and autonomic fibers to the viscera and skin, the posterior root supplies sensory fibers to the corresponding dermatome as well as to the muscles and viscera In mammals, as a result of the outgrowth of the embryonic limbs the orderly arrangement of the metameres from before backwards becomes altered In the body of the adult, the primitive meta-

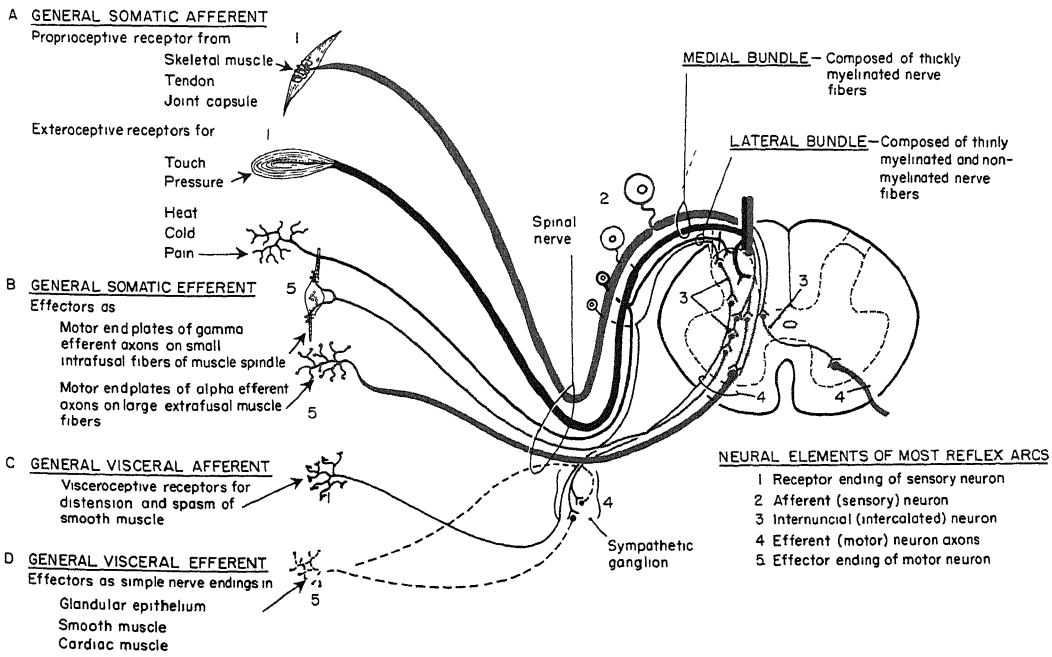


Fig 9 22 Diagram of functional components of a thoracic spinal nerve showing distribution of dorsal root fibers as they enter the spinal cord Skeletal muscle afferent and efferent fibers are shown in brown Numbers identify elements as listed at lower right From Truex and Carpenter 1969

meric disposition is observed only in the trunk An apparent segmentation is present, nevertheless, in the spinal nerves because they exit between pairs of vertebrae But then as their fibers take their courses to or from the limbs, they form the brachial and lumbosacral plexuses and, after intermingling freely, issue again as the peripheral nerves The latter, in consequence, are composed of fibers derived from two or more spinal segments, and fibers from a given segment pass into several peripheral nerves The muscles supplied by a given spinal segment do not necessarily lie adjacent to one another (the coracobrachialis, for example, is innervated by the same segments as those which supply the muscles of the thumb) and a single muscle may derive its nerve supply from more than one spinal segment As development proceeds and the limbs grow out from the trunk, the dermatomes become arranged in a series of narrow areas lying for the most part in the long axis of the limb (Fig 9 23) The skin and muscles of the limbs also tend to move away from the visceral structures with which they were originally associated and, in the adult, structures innervated by a common spinal segment may be widely separated Thus, the diaphragm is innervated (through the phrenic) from the 3, 4, and 5 cervical segments which also supply skin and muscle in the region of the neck and shoulder The heart receives sensory and autonomic fibers from the upper thoracic segments, these segments also supply sensory fibers to the skin over the inner aspect of the arm and hand and upper part of the thorax This topic will be considered further in its relation to the phenomenon known as referred pain

Several methods have been employed by different investigators in mapping out the dermatomes in animals and in the human subject The *anatomical* method is laborious and consists in tracing the fibers of a spinal root to their terminations in the skin A *physiological* method ("isolation" or "sensory remainder" method) was employed by Sherrington in cats and monkeys The area of skin supplied by a given segment was demarcated by dividing at least two of the sensory roots above and two more below it The region supplied by the uncut root was revealed as a sensitive area of skin bounded above and below by an anesthetic zone supplied by the cut fibers (No anesthesia follows cutting of the dorsal root fibers of a single spinal nerve, because of overlap from adjacent segments) Foerster (1933) used similar methods to map out the dermatomes in man He also stimulated the posterior roots at operation, and used the resulting vasodilatation as a means of demarcating dermatomes Similar maps were prepared by Head (1893) from studies of cases of herpes zoster, a condition due to a viral infection of cells, usually in a single dorsal root ganglion

The Tracts of the Spinal Cord and Brainstem

Fiber tracts of the cord are divisible into two main groups *a*) *Long tracts*, both ascending and descending, connect the cord with other parts of the central nervous system *b*) *Short tracts* (intersegmental or association tracts, ground bundles) begin and end within the spinal cord and connect different segments The fiber tracts making up the substance of the respective funiculi are shown diagrammactically in Figure 9 24

SUMMARY OF FUNCTIONS OF PRINCIPAL ASCENDING TRACTS

A) Fibers Mediating Conscious Sensations. *1*) Fibers mediating thermal and painful sensations are part of the lateral divisions of the posterior nerve roots, and enter *Lissauer's*

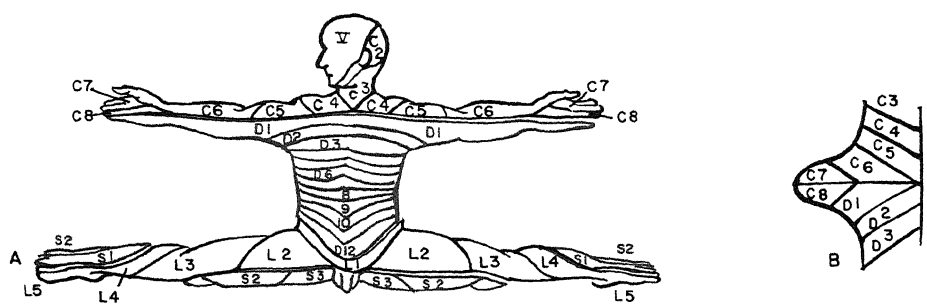


Fig 9 23 *A* Levels of the spinal cord from which various areas of the surface of the body receive sensory innervation *B*, Diagram to show how development of limb bud alters dermatomal representation

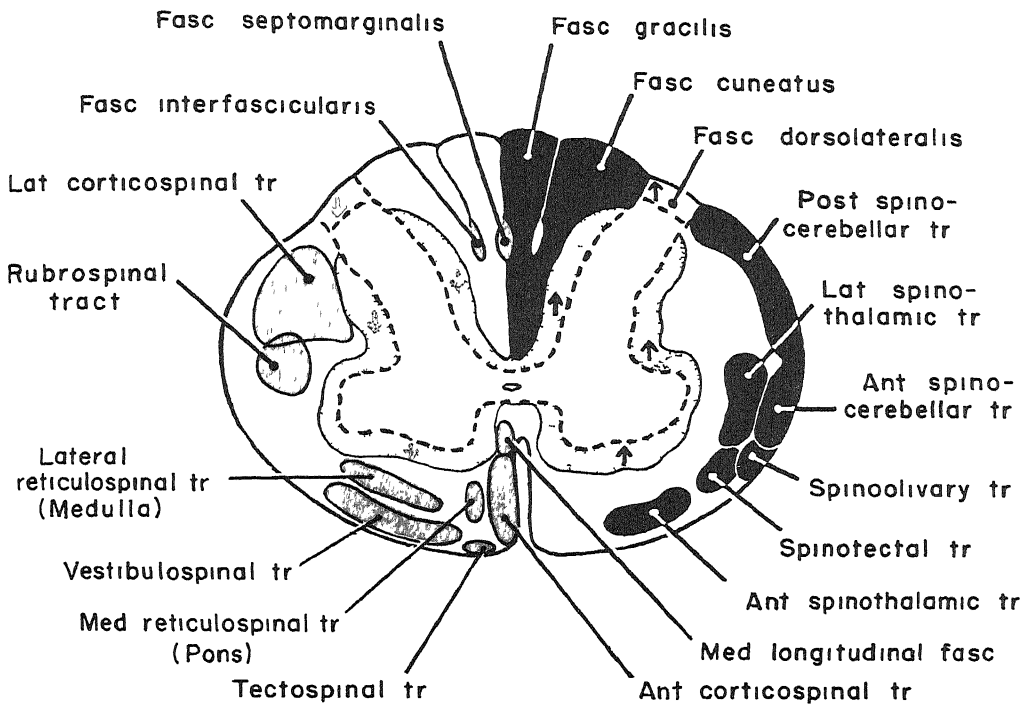


Fig 9 24 Location of major spinal pathways Ascending pathways are shown on the right in black descending paths are on the left in brown The fasciculus proprius system (*stippled*) and dorsolateral fasciculus contain both ascending and descending fibers From Truex and Carpenter 1969

tract of the spinal cord Immediately, or after a very short upward or downward course, they enter the dorsal horn and connect with second order neurons having axons that cross to the opposite side in the white commissures and ascend as the *ventrolateral spinothalamic tract* This tract occupies a position in the brainstem lateral to the medial lemniscus and ends in the posterolateral part of the ventral nucleus of the thalamus (Fig 9 25)

2) The fibers conveying all other sensory impulses, e g , from muscles and joints (sense of movement and position), and those mediating the senses of vibration, touch, pressure, tactile localization, and spatial discrimination enter the cord in the medial division of the posterior roots and ascend in the *fasciculus gracilis* and *fasciculus cuneatus* The fibers mediating all of these sensations (except for some in the ventral spinothalamic tract concerned with touch and probably with tactile localization) pass without crossing to the nuclei of these fasciculi in the medulla oblongata, where they connect with secondary neurons The axons of the latter constitute the *internal arcuate fibers*, they decussate with those of the opposite side and ascend to the thalamus as the *medial lemniscus* Tertiary neurons convey the impulses from thalamus to cerebral cortex

B. Sensory Pathways That Do Not Reach the Cortex Impulses from receptors in muscles, tendons, and joints, as well as some from tactile receptors that are not destined to reach the thalamus and cerebral cortex are conveyed into the spinal cord by fibers composing the medial divisions of the posterior roots These fibers connect at once with cells of *Clarke's column* and are continued upwards in the *dorsal* (direct) and *ventral* (indirect) *cerebellar tracts* of the same side, mainly, but also of the opposite side Some of these impulses, viz., those which enter the cervical cord, are conveyed by fibers of the cuneate fasciculus to the cuneate nucleus, to be relayed by secondary neurons (*external arcuate fibers*) via the inferior peduncles to the cerebellum (Fig 9 25)

SENSORY DISABILITIES FOLLOWING LESIONS

In a *transverse spinal lesion* interrupting solely the posterior fasciculi, vibratory and position senses are lost but the sensation of touch is retained, since the fibers which have entered the gray matter and crossed below the level of the lesion to ascend in the ventral spinothalamic tract have escaped injury Like-

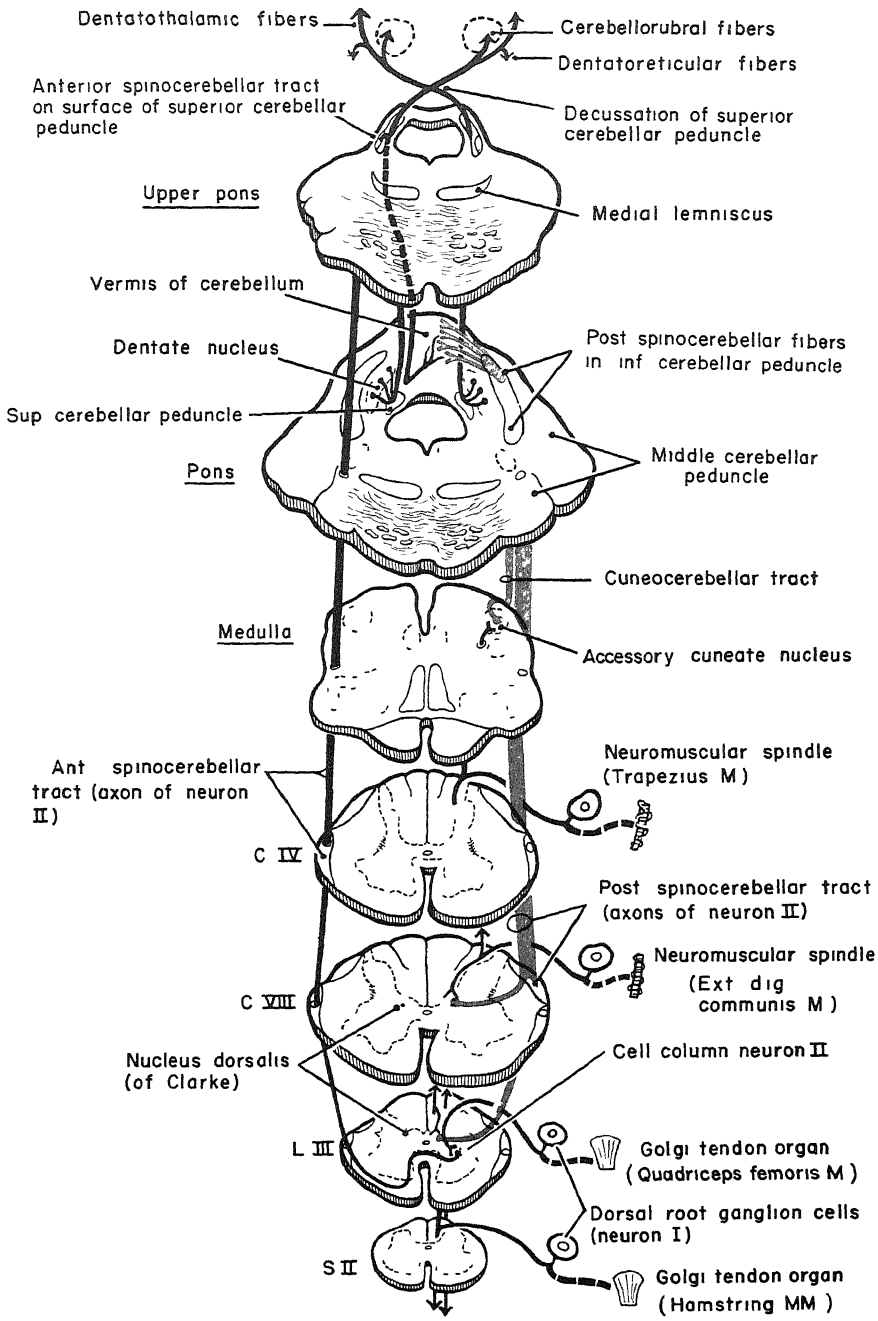


Fig 9 25 Proprioceptive pathways ascending from spinal cord through brainstem with connections in cerebellum pons and on to thalamus Spinal levels are indicated on the left From Truex and Carpenter 1969

wise, although a *hemisection of the spinal cord* destroys vibratory and position senses on the same side and pain and temperature sensibilities on the opposite side it does not abolish tactile sensibility as this is conveyed in both crossed and uncrossed pathways

A lesion in the *lower part of the brainstem* may involve one of the sensory pathways exclusively of the others Thus, an injury localized to the outer part of the lower pons or of the medulla oblongata by injuring the spinal lemniscus may cause loss of sensation to pain,

heat, and cold over the *opposite half* of the body and leave muscle sense and tactile discrimination intact Sensory loss of this nature accompanied by ipsilateral cerebellar symptoms occurs as a result of the occlusion (as by thrombosis or embolism) of the *posterior inferior cerebellar artery* Usually also, as a result of the involvement of the bulbospinal tract of the trigeminal nerve the face on the same side as the occluded vessel shows a loss of pain and temperature sensibilities whereas tactile sensation which is transmitted by the main sensory nucleus is largely unaffected A lesion more centrally placed may, by implicating the medial lemniscus alone, cause the converse type of dissociated sensory defect, namely, loss of the sense of position of the limbs and of spatial discrimination with retention of sensibility to pain, heat, and cold In lesions at higher levels in the brainstem all forms of sensation are likely to be involved more or less equally

DESCENDING TRACTS OF SPINAL CORD AND BRAINSTEM

1) Corticospinal (Cerebrospinal) Tracts. From their origins in the cerebral cortex these tracts descend through the cerebrum, through the internal capsule on each side, the crus cerebri of the mesencephalon, the pons of the metencephalon, and the pyramid on the inferior surface of each side of the medulla oblongata (Fig 9 26) At the lower border of the medulla each tract divides into two bundles of unequal size, the larger of which crosses to the opposite side and descends in the posterior part of the lateral funiculus of the cord as the *lateral corticospinal tract* The remaining fibers uncrossed, descend in the anterior funiculus as the *anterior corticospinal tract* In most instances the direct or anterior corticospinal tract is but a small part of the total number of corticospinal fibers (usually not over 10%), but the number of fibers in the anterior tract varies and in rare instances most of the corticospinal fibers lie in this position In most cases these anteriorly placed fibers descend only as far as the upper thoracic spinal cord Even these fibers cross to the opposite side before terminating on the ganglion cells of the anterior gray column There are also direct corticospinal fibers lying in the lateral columns of the spinal cord which do not cross to the opposite side Through these the precentral motor cortex sends impulses to ipsilateral skeletal muscles

All corticospinal fibers whether crossed or uncrossed connect with the large motor cells of the ventral gray columns (anterior horn cells) The connections are of two types, *a*) direct synapses (20 to 30%) with the motor neurons, *b*) indirect connections with these cells through an internuncial neuron whose cell body is also in communication with a posterior root fiber on each side It has been estimated that from 75 to 90% of the corticospinal fibers terminate in the cervical (55%) and thoracic (20%) regions of the spinal cord, and about 25% in the lumbar and sacral regions

For many years an erroneous belief was entertained that all of the fibers in the pyramids on the ventral surface of the medulla oblongata arose from the Betz cells of the precentral gyrus and were responsible for all voluntary activity of the skeletal musculature It is now known that this concept contains many errors Only a very small percentage of the corticospinal fibers arise from the gigantic Betz cells of the precentral gyrus, approximately 3 to 4% The remainder of the corticospinal fibers arise from other cells in the precentral gyrus, from cells in the parietal cortex and perhaps other areas There is a question as to whether all of the fibers in the medullary pyramids are of cortical origin, and some investigators have even questioned whether they are all descending fibers Certainly, all voluntary muscular activity is not achieved exclusively by corticospinal fibers nor by fibers contained in the medullary pyramids There are multisynaptic fiber pathways from the cortex down to the anterior horn cells of the spinal cord, which are also capable of producing voluntary movement

2) Corticobulbar Fibers. The fibers passing from the lower part of the motor area of the cortex to the cranial nuclei constitute the important *corticobulbar tract* These bear precisely the same relationship to the cells of the cranial motor nuclei as does the corticospinal tract to the anterior horn cells of the cord The corticobulbar fibers, however, cross at various levels throughout the brainstem The location of these fibers and the corticospinal systems in the internal capsule is considered in the following section on the telencephalon

3) Vestibulospinal Tracts. Two tracts, a ventral and a lateral, are commonly desig-

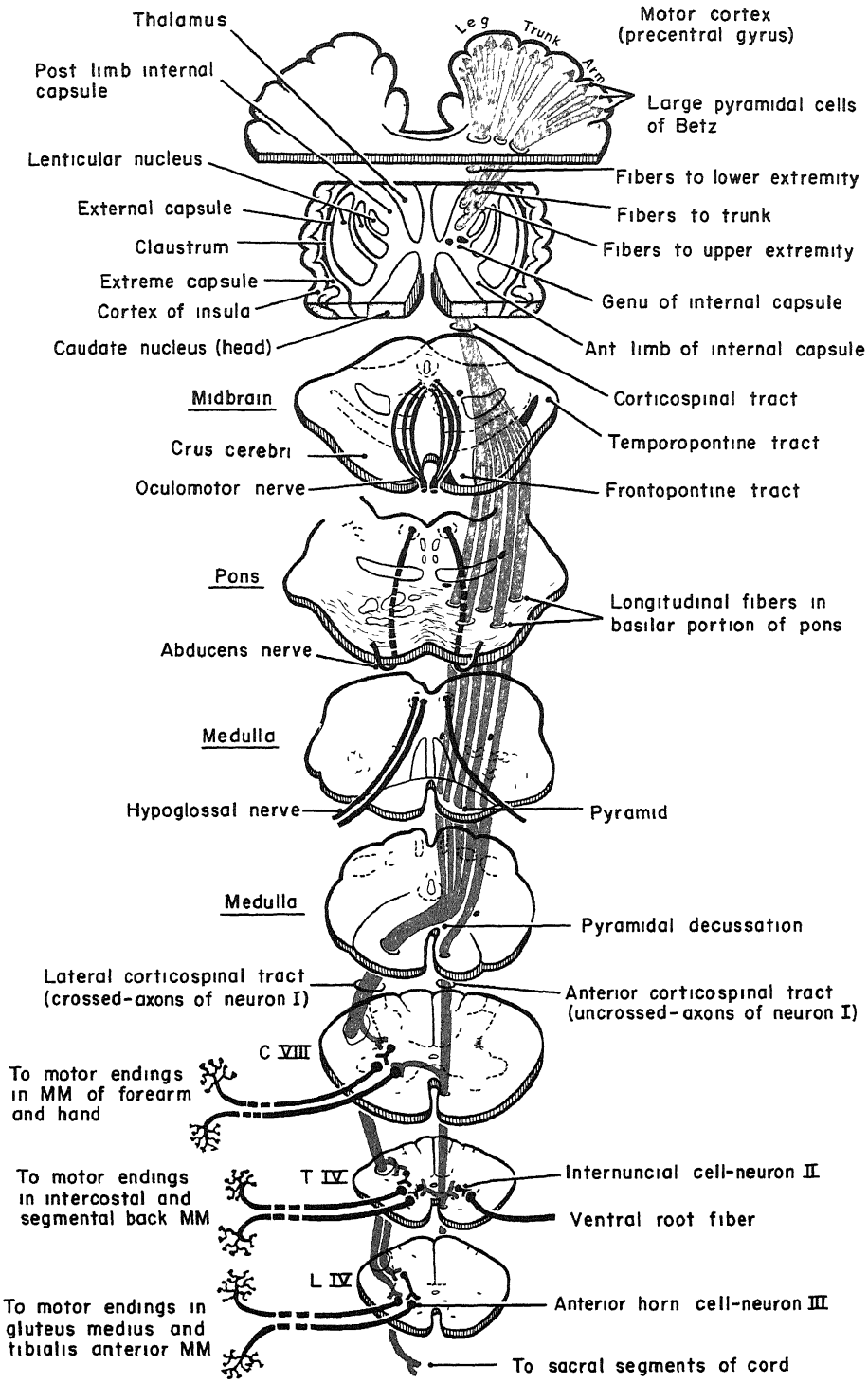


Fig 9 26 Diagram showing course of corticospinal fibers in lateral and anterior corticospinal tracts beginning in pyramidal cells of cortex running through internal capsule basis pedunculi pyramid and into spinal cord Location of corticobulbar tracts is shown also (*right side*) From Truex and Carpenter 1969

nated, although there is no evidence that they serve different functions or that there is any significant difference between them The *ventral vestibulospinal tract* arises predominantly from the spinal or inferior vestibular nucleus (in the cat) and descends in the forepart of the anterior funiculus of the spinal cord just lateral to the anterior corticospinal tract It is a downward extension of the medial longitudinal fasciculus (see below, p 9–30) The *lateral vestibulospinal tract* is a prominent descending pathway lying in the anterolateral part of the spinal cord It arises from the lateral vestibular or Deiter's nucleus and extends throughout the entire length of the spinal cord Most of its fibers are uncrossed although a few arise from the contralateral lateral vestibular nucleus The vestibulospinal fibers terminate either upon the anterior horn cells or upon small internuncial neurons which in turn form synapses with the motor cells of the anterior gray horns of the spinal cord

4) **Tectospinal Tracts** of each spinal half arise from the contralateral superior and inferior colliculi and descend at first as a single compact bundle through the reticular formation of the brainstem Upon reaching the cord the fibers are segregated into two fasciculi, a ventral and a dorsal, which descend in the ventral and lateral white columns, respectively The fibers synapse with the spinal motoneurons, either directly or through internuncial neurons, especially in the cervical region, wherein are situated the centers for the neck muscles Impulses from the retina are received by the superior colliculus which through such impulses, and its spinal connections serves as a center for the integration of visual impressions with body movement, especially of the head (*visuospinal reflexes*) This center is probably of less importance in man, with his highly developed visual cortex, than in animals Auditory reflexes may also be mediated by this tract and the inferior colliculus

5) **Rubrospinal Tract** arises from the large cells (nucleus magnocellularis) in the posterior part of the red nucleus The fibers immediately upon leaving the red nucleus cross to the opposite side (*Forel's decussation*) and descend through the reticular formation of the pons and medulla to enter the lateral funiculus of the cord ventral to the lateral corticospinal tract In carnivorous animals, e g , cat and dog, the rubrospinal tract is large and important In the primates, particularly in man, the rubrospinal tract is so small that its existence has been questioned, at any rate it has little if any known functional significance

6) **Olivospinal (Bulbospinal) Tract of Helweg** Its fibers are presumed to arise from cells in the neighborhood of the inferior olivary nucleus and descend in the ventral and lateral part of the lateral funiculus However, its existence has been seriously questioned It descends no farther than the cervical region The fibers synapse with motor neurons The functions of this tract are unknown It and the thalamo-olivary tract constitute a possible path whereby impulses from the thalamus may reach the spinal centers

7) **Reticulospinal Tracts** arise from cells scattered through the reticular formation of the brainstem They are two in number, one in which crossed fibers predominate descends in the anterior funiculus, the other composed mainly of uncrossed fibers, in the lateral funiculus They connect with the motor neurons in the anterior horns and probably with cells in the lateral horn Impulses conveyed over this complex pathway are concerned with the regulation of the reflux activity of the skeletal musculature, the control of the vegetative nervous system, adjustment of the threshold of peripheral end-organs such as the muscle spindles and of the sensory pathways

INTERSEGMENTAL TRACTS OF THE SPINAL CORD

There is in each funiculus of the cord lying close to the gray matter an intersegmental fasciculus or ground bundle These, as already mentioned, serve to link spinal segments of different levels Their fibers arise from cells in the gray matter and after an ascending or descending course of variable length end around cells of the same or of the opposite side at a higher or a lower level They include groups of fibers that have the following names *lateral intersegmental fasciculus* or *lateral ground bundle*, the *anterior intersegmental* or *sulcomarginal fasciculus, septomarginal fasciculus,* and the *posterior intersegmental fasciculus*

THE LONGITUDINAL FASCICULUS (OR BUNDLE)

The medial (posterior) longitudinal fasciculus (or bundle) is a tract of great physiolog-

ical importance and is present in all vertebrates It is composed of fibers which connect the 3rd, 4th, and 6th cranial nerves with one another and with the vestibular nuclei, and makes connections with the motoneurons of the cervical segments of the spinal cord The tract lies near the midline and extends from the cervical segments of the spinal cord, where it is continuous with the anterior intersegmental fasciculus (anterior ground bundle) to the floor of the 3rd ventricle In the medulla oblongata it lies immediately subjacent to the floor of the 4th ventricle, in the pons it courses through the formatio reticularis, and in the midbrain, lies in relation to the gray matter of the floor of the Sylvian (cerebral) aqueduct It also received fibers from the lateral lemniscus (auditory path) and from the superior colliculus, through which it is in communication with the optic pathway The essential function of the medial longitudinal bundle is the coordination of reflex movements of the ocular and neck muscles in response to labyrinthine, auditory, and visual stimuli

STRUCTURE OF THE BRAINSTEM

At lower levels of the medulla oblongata the homology of brainstem and spinal cord is obvious upon examination of cross sections As sections are cut progressively more rostrally, however, the brainstem exhibits more and more features that obscure its similarity to the cord, and one has to look closely to distinguish familiar landmarks As noted already, the principal ascending and descending pathways of the spinal cord pass through or on the brainstem, these provide useful clues as to the homology Other clues are found in the sensory and motor nuclei of the cranial nerves If the ventricular system of the brainstem is regarded as a modified, displaced, and flattened out continuation of the central canal of the spinal cord, the gray matter adjacent to the ventricles and aqueduct is then seen to be continuous with the gray columns of the spinal cord Instead of the dorsal-ventral relation found in the cord, however, the gray matter in the floor of the 4th ventricle possesses sensory elements laterally, autonomic motor components in an intermediate position, and motor cells medially This comes about because the dorsal tips of the gray columns have been, in effect, separated until the H of the spinal cord has become an open V (Fig 9 27) In many instances even this relationship is obscured by development of other structures, with a result that motor nuclei that theoretically should lie in or near the central gray substance near the midline are found in what appears to be a remote position, as in the case of certain nuclei of the facial nerve and of nerves IX, X, and XI, as described later

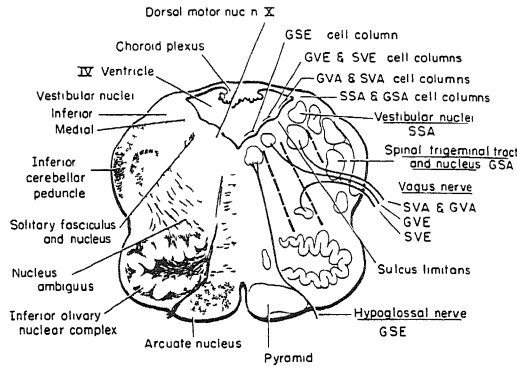


Fig 9 27 Schematic cross section of medulla oblongata to show appearance in stained section (*left*) and pattern of organization (*right*) Cell columns that contain nuclei of origin and termination of cranial nerves are abbreviated as follows *GSA* general somatic afferent *GSE* general somatic efferent *GVE* general visceral efferent *SVA* special visceral afferent *SVE* special visceral efferent From Truex and Carpenter 1969

Most of the apparent distortion of relationships in the brainstem, however, is associated with development of the cerebellum, its related structures, and its connections The principal functional relation of the cerebellum is with the cerebral cortex, their connection is achieved through fibers from the cerebrum that pass via the internal capsule and cerebral peduncles to nuclei imbedded in the pons, and thence by fibers of the pons that enter the cerebellum via the middle cerebellar peduncle For this system the main path of exit from the cerebellar nuclei is the brachium conjunctivum that runs in the rostral or superior cerebellar peduncle to the midbrain and thalamus It is the fact that both of these pathways contain a relatively large number of fibers, plus the cells located in the pons, that give the level of the metencephalon those characteristics that make it seem so different from the medulla below it and the mesencephalon above

That portion of the brainstem that seems to run continuously from spinal cord through medulla oblongata, metencephalon and mesencephalon (Fig 9 20) is known as the *teg-*

mentum (The name comes from a word that means "cover," and it is true that generally the tegmentum has a dorsad position, especially at the level of the pons In the mesencephalon, however, it in turn is covered by the corpora quadrigemina or *tectum*—a word that means "roof") The tegmentum contains the complex of cells and fibers known as the *reticular formation*, and certain of the nuclei of the cranial nerves In the medulla it includes all structures except the inferior cerebellar peduncles above, and the complex of olivary nuclei and the pyramids lying inferiorly The "olives," because of their conspicuous projection to the cerebellum, may be classified with other structures of the brainstem related to the development of the cerebellum

At the level of the metencephalon, a level usually but inexactly known as the pons, the tegmentum is hidden by the cerebellum above it and the pons below (Fig 9 20) At the level of the mesencephalon it is the corpora quadrigemina that lie in the superior position, with the substantia nigra and crus cerebri placed inferior to the tegmentum on each side The *red nuclei* are prominent collections of neurons located ventromedially in the tegmentum of the mesencephalon, surrounded by fibers of the brachium conjunctivum that seem to form a "capsule" for the nucleus

The Reticular Formation

The central portion of the tegmentum of the brainstem is occupied by a diffuse ill-defined mass of nerve cells and fibers known as the *reticular formation* It extends downward into the spinal cord and upward into the diencephalon It contains identifiable cell groups, however, as these have not been shown to have functional significance no cognizance of these nuclei will be taken here The reticular formation has been defined (see Segundo, 1956) as including all areas within the brainstem (excepting primary afferent pathways) which when stimulated will produce "arousal" (chap 8, p 9–150) This includes the bulbar reticular formation, the pontomesencephalic tegmentum, the hypothalamus, the subthalamus, and the ventromedial portion (including the intralaminar nuclei) and the reticular nucleus of the thalamus

The reticular formation receives afferent impulses from a great variety of sources These connections are by means of both direct afferent fibers and collaterals from the main afferent pathways to the cerebellum and thalamus Thus, the reticular formation receives afferent impulses from the optic, the olfactory, the auditory, and the gustatory systems as well as from the spinal and trigeminal pathways which bear impulses concerned with tactile, painful, thermal, and vibratory sensibilities and sensations from the muscles, joints, and tendons, and afferent impulses from the viscera and other internal structures of the body In addition afferent fibers converge on the reticular formation from the cerebellum, the basal ganglia, the thalamus, the rhinencephalon and the cerebral cortex Peculiarly, only certain specific portions of the cerebral neocortex seem to send afferent impulses to the reticular formation These are the orbital, oculomotor, sensorimotor and posterior parietal cortices, the superior temporal and cingular gyri, and the temporal pole In addition the activity of the reticular formation is influenced by certain hormonal and chemical changes (adrenalin, acetylcholine, carbon dioxide, etc) and by many drugs (barbiturates, anesthetics, tranquilizing compounds, etc)

In turn, efferent impulses from the reticular formation are conducted *i) cephalad* into the brain, particularly to all parts of the cerebral cortex, and *ii) caudad* into the spinal cord Because of the multisynaptic character of the pathways to and from the reticular formation and the small caliber of many of the fibers, reticulocortical conduction is much slower than in the primary sensory system The reticulospinal fibers run bilaterally in the lateral and ventral white columns of the spinal cord and transmission in this system is thought to be quite rapid

Cranial Nerves

Cranial nerves are numbered from above down, and will be considered in that order here even though this is contrary to the outline of the other material in this chapter (Figs 9 28 and 9 29) The nuclei of origin of the motor components of the cranial nerves will be classified as follows

1) Somatic efferent—to striated muscle via nerves III, IV, VI, and XII,

2) General visceral efferent—to smooth muscle and glands, via nerves III, VII, IX, and X,

3) Special visceral efferent—to striated

muscles derived from branchial arches, fibers run in nerves V, VII, IX, and X (and XI)

The nuclei of termination of afferent fibers, the nuclei that contain the cell bodies of second order sensory neurons, have the following classification

4) *Visceral afferent*—receiving taste and other fibers brought via nerves VII, IX, and X,

5) *General somatic afferent*—sensory nucleus and nucleus of the spinal tract of nerve V, a small contribution from nerve X,

6) *Special somatic afferent*—vestibular and cochlear nuclei of nerve VIII

Autonomic motor components of the cranial nerves belong to what is known as the parasympathetic portion of that system Their significance is discussed in the following chapter The general location of all of these nuclei is represented in relation to surface anatomy of the brainstem in Figure 9 30

OLFACTORY AND OPTIC NERVES (I AND II)

As noted earlier, these are not nerves like the other 10, in that they develop differently and have different types of connections The afferent fibers they contain do not arise from cells located in ganglia outside the central nervous system Rather, these two nerves should be regarded as fiber tracts of the brain Their function is considered in section 8 The olfactory nerve seems to contain only afferent fibers, whereas the optic nerve includes also fibers passing from the central nervous system to the retina

OCULOMOTOR, TROCHLEAR AND ABDUCENS NERVES (III IV AND VI)

The oculomotor nerve innervates all extrinsic muscles of the eye except the superior oblique muscle (innervated by trochlear nerve) and lateral rectus (by abducens nerve) These motor fibers arise from three separate somatic motor nuclei that belong to the column that includes also the motor nucleus of the hypoglossal nerve Autonomic fibers (preganglionic, see following chapter) to intrinsic muscles of the eye also run with the oculomotor nerve from their origin in the Edinger Westphal nucleus in the mesencephalon—the most rostral

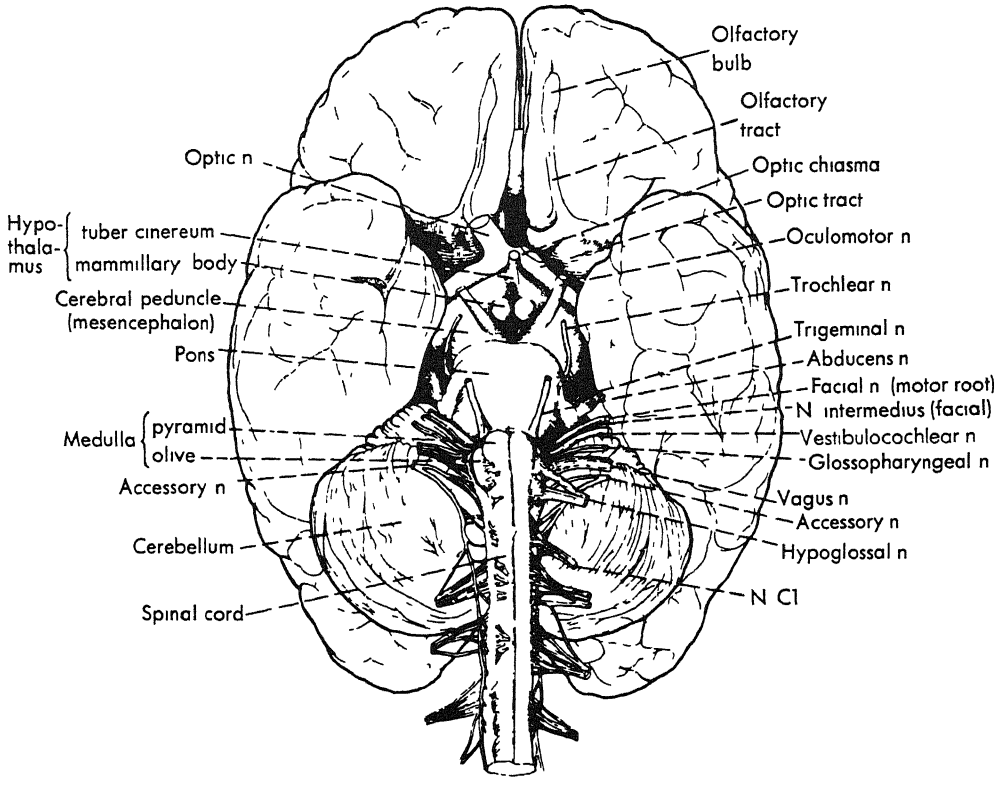


Fig 9 28 Inferior surface of human brain and upper part of the spinal cord From Henle 1868 reproduced from Hollinshead, 1967

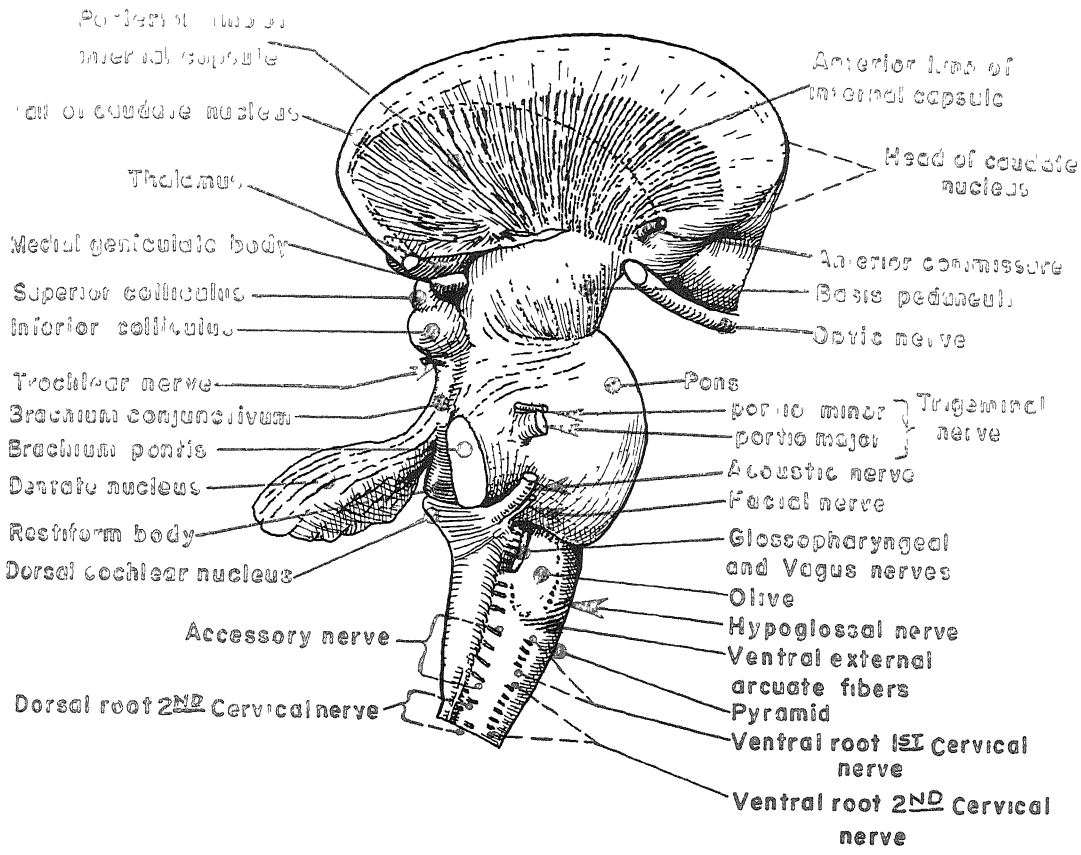


Fig 9 29 Lateral view of human brainstem Cranial nerves have been sectioned cerebellum has been removed by section of the three peduncles on each side all portions of the cerebrum lateral and superior to the internal capsule have been dissected away (Ranson and Clark 1959)

of the four general visceral motor nuclei Afferent fibers from extrinsic muscles of the eye are said to run in all three of these nerves, but their central connections are uncertain

The oculomotor nerves originate in the mesencephalon, and leave the brainstem inferiorly and medially on each side between the midline and the medial border of the cerebral peduncle Fibers of the trochlear nerve run dorsally around the central grey substance of the mesencephalon, cross in the region superior to and between the inferior colliculi, leave the brainstem, and then course around it en route to the orbit By its crossing this nerve becomes unique among nerves supplying extraocular muscles The abducens nerve makes its exit from the brainstem at the caudal border of the pons not far from the midline

TRIGEMINAL PATHWAYS

The trigeminal nerve appears on the lateral surface of the pons The nerve has three roots, *i*) a large *sensory*, *ii*) a *mesencephalic* that is also sensory, and *iii*) a small *motor* (Fig 9 31)

The fibers of the *sensory root* convey impulses from the anterior part of the scalp, from the skin of the forehead and face, with the exception of an area over the lower border of the mandible It also supplies the mucous membrane of the mouth, anterior two-thirds of tongue and nose, the cornea and conjuctivae, and part of the intracranial dura mater This root arises from the cells of the gasserian (semilunar) ganglion, which corresponds to a spinal posterior root ganglion The processes of the ganglion cells divide into peripheral and central branches The former enter into the composition of the peripheral divisions (ophthalmic, maxillary, and mandibular) of the nerve, the central branches (which constitute the sensory root of the nerve) enter the pons in close association with the fibers of the motor root and divide into an ascending and a descending group in a manner homologous with a spinal root fiber The *ascending fibers* convey impulses of light touch, tactile discrimination and localization, and of the sense of position and passive movement They end in a

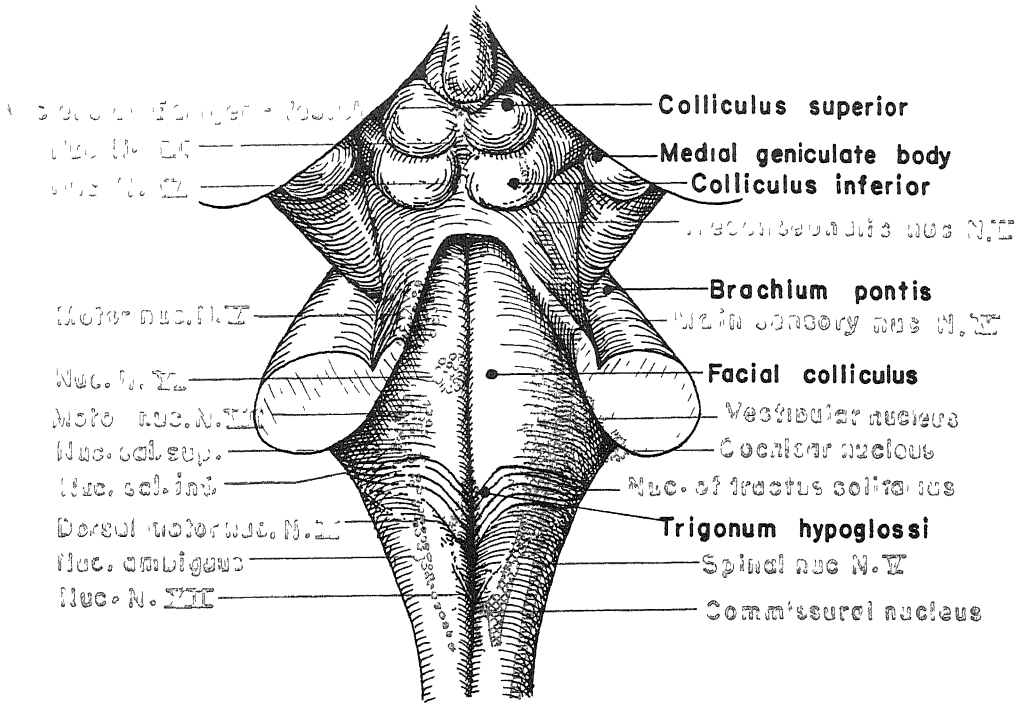


Fig 9 30 Dorsal view of human brainstem with positions of cranial nerve nuclei projected upon the surface Motor nuclei are on the left sensory nuclei on the right Herrick from Ranson and Clark 1959

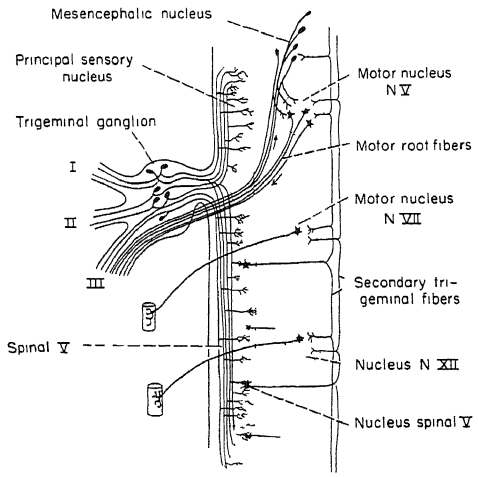


Fig 9 31 Diagram of peripheral and central distribution of sensory and motor fibers in trigeminal nerve I ophthalmic II, maxillary, and III mandibular divisions From Truex and Carpenter 1969

nucleus (*main* or *upper sensory nucleus of the trigeminal nerve*) situated in the pons, deep and lateral to the motor nucleus The axons of cells in this nucleus ascend with those from the spinal nucleus of the trigeminal nerve (see below) to form the *trigeminal lemniscus* The *descending fibers* constitute the *bulbospinal tract of the trigeminal nerve* They subserve sensations of pain and temperature over the entire trigeminal area They descend through the pons and medulla oblongata and may be traced as far as the 2nd cervical segment This tract dwindles gradually in its descent, its fibers terminating around cells in the substantia gelatinosa of Rolando which in this situation is referred to as the *spinal nucleus of the trigeminal nerve* From the spinal nucleus the axons of secondary neurons swing downward in the spinal tract of the trigeminal nerve as far as the upper cervical spinal cord They then cross the midline and turn upward In the upper part of the medulla oblongata they join fibers arising from the upper sensory nucleus to form the *trigeminal lemniscus* (or fillet) The trigeminal lemniscus lies in close association with the medial and the spinal lemnisci (spinothalamic tracts) (p 9–25), its fibers terminate in the posteroventral nucleus of the thalamus

The main sensory nucleus of the trigeminal corresponds to the nuclei of the posterior columns of the spinal cord (nuclei gracilis and cuneatus) and like the latter sends fibers to the cerebellum through the inferior cerebellar peduncle The spinal tract of the trigeminal

nerve may be looked upon as corresponding to the tract of Lissauer (p 9–24), and the spinal nucleus as an extension upwards of the substantia gelantinosa of Rolando

A number of reflexes are mediated through the afferent fibers of the trigeminal nerve, e g , corneal, winking, sneezing, and oculocardiac

The *mesencephalic root* consists of a small bundle of afferent fibers which run in company with the fibers of the motor root Entering the pons they ascend to the *mesencephalic nucleus* of the trigeminal, an elongated collection of first order sensory nerve cells extending from the level of the motor nucleus to the upper region of the midbrain The nucleus is looked upon as a group of cells homologous with the gasserian (semilunar) ganglion and the posterior root ganglia of the spinal nerves, but which has migrated or been "drawn" into the brainstem at an early period of phylogenetic development It is believed to receive proprioceptive impulses from the muscles of mastication and possibly from the ocular muscles

The *motor root* fibers arise from a nucleus in the upper part of the pons underlying the lateral part of the floor of the 4th ventricle The nucleus receives fibers from the corticobulbar tract of the opposite side and also probably from the same side The motor root after its emergence from the brainstem, travels peripherally with the sensory root and, passing deep to the gasserian ganglion, joins the mandibular (3rd) division of the trigeminal nerve to supply the muscles of mastication (temporal, masseter and pterygoids)

Lesions Involving the Trigeminal Pathways Either the peripheral portions or the central connections of the nerve may be the seat of disease A lesion of the nerve peripheral to the ganglion is more likely to involve only one of its three divisions Pain, or loss of sensibility over the distribution of the divison affected, may result If the gasserian ganglion is involved the disturbance of sensation usually involves the area of distribution of more than one division, often of the entire nerve If as a result the eye is rendered anesthetic the patient may develop an ulceration of the cornea and an inflammation of the cornea and conjunctiva This has often been referred to as a neuroparalytic keratitis and has been attributed to the loss of some obscure trophic influence of the nerve Actually it results from the fact that the eye is anesthetic and can be injured easily by a foreign body or by scratching without any discomfort to the patient The eye is then likely to be neglected by the patient because it is not painful As a result he may lose his sight or even his eye

Herpetic eruptions may result from the involvement of the gasserian ganglion Spontaneous herpes zoster usually involves the area of distribution of the ophthalmic division and is often serious because the lesions on the cornea may lead to corneal opacity and loss of vision Such herpes is also commonly followed by a very distressing postherpetic neuralgia which is very difficult to treat satisfactorily

The trigeminal nerve, or one of its divisions, is sometimes the seat of a severe and intractable type of pain which recurs in paroxysms (*trigeminal neuralgia, tic douloreux*) The cause of the affection is unknown

Paralysis of the muscles of mastication with reduction in the strength of the bite on the affected side and deviation of the mandible toward the paralyzed side when it is opened occurs when the motor division is involved

THE FACIAL NERVE

The facial nerve consists of a large *motor* and a small *sensory* portion The two portions or roots appear at the lower border of the pons and enter the internal auditory meatus in company with the auditory nerve (Fig 9 32)

The *sensory* root of the facial, which is also known as the *nervus intermedius of Wrisberg*, contains not only afferent fibers but secretory and vasodilator (parasympathetic) fibers as well

The *sensory fibers* arise from the cells of the geniculate ganglion The peripheral processes of these cells are distributed through the chorda tympani branch of the facial nerve to the taste buds and mucous membrane of the anterior two-thirds of the tongue, and through the nerve to the pterygoid canal (formed by the union of the greater superficial petrosal and the deep petrosal nerves), and the sphenopalatine ganglion to the lacrimal gland and the mucosa of the soft palate and posterior part of the nose The chorda tympani branch joins the trunk of the lingual nerve through which it is conveyed to the floor of the mouth The central processes of the ganglion cells end in the *sensory nucleus* situated in the upper part of the *tractus solitarius* Fibers arise from the

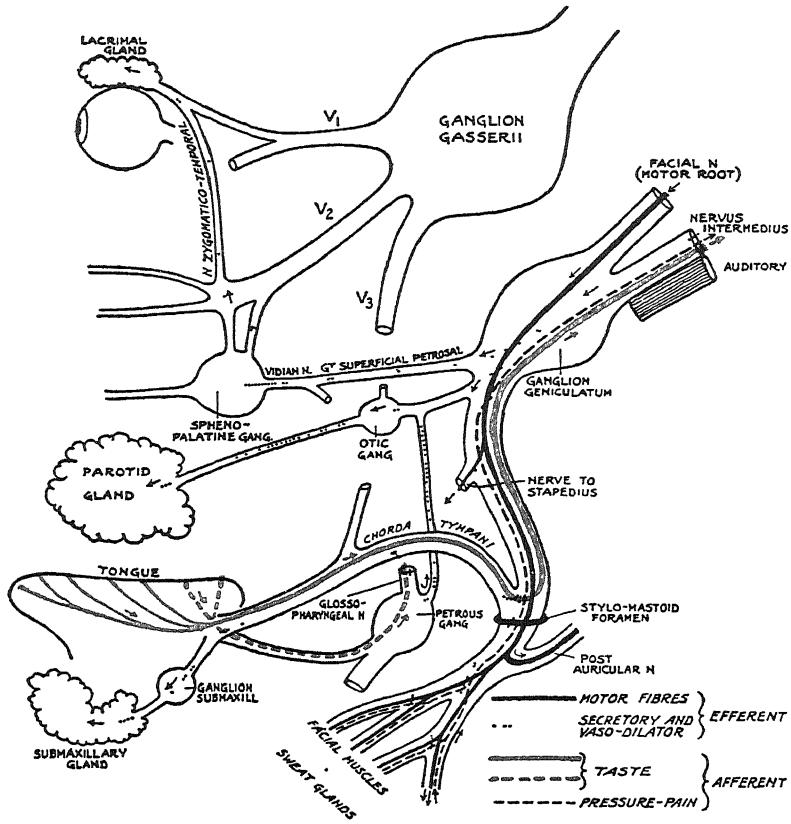


Fig 9 32 Components and distribution of the facial nerve (Treux and Carpenter 1969)

latter and ascend in the medial fillet of the opposite side to reach the thalamus

The *parasympathetic fibers* arise from the *superior salivatory nucleus* which lies in close relation to the motor nucleus of the facial After leaving the brainstem in the sensory root the parasympathetic secretory and vasodilator fibers pass via the *great superficial petrosal* nerve and *nerve to pterygoid canal* (formed by the union of the great superficial and deep petrosal nerves) to the *sphenopalatine ganglion* from where the postganglionic fibers arise to pass to the lacrymal gland, and to the vessels and glands of the palate and posterior part of the nose The secretory and vasodilator fibers to the submaxillary and sublingual glands leave the facial with the taste fibers in the chorda tympani branch

The *motor part* of the facial is conveyed through the facial canal of the temporal bone to the stylomastoid foramen After its emergence from the latter it is distributed to muscles of the face, auricle, and forehead. The motor fibers arise from a nucleus in the lower part of the pons and pass backwards to the lower end of the nucleus of the abducent nerve Then ascending to the upper end of the latter they bend and sweep downwards and forward to where they leave the brain As the facial fibers take this arched course to the point of their emergence from the brainstem, they, together with the abducent nucleus, form a prominence in the floor of the 4th ventricle, known as the *facial colliculus*

Facial Paralysis. The effects of interruption of the facial pathway vary in certain important features according to the level at which the injury occurs In the paralysis resulting from division to the trunk of the facial nerve all the muscles of the affected side of the face are completely paralyzed The subject is unable to close the eye owing to paralysis of the orbicularis oculi, or to frown, the eyebrow droops The mouth is drawn over to the sound side by the unparalyzed muscles and the muscles of the affected side do not take part in the facial expression of emotional states, e g, laughing or crying

Facial paralysis may result from a lesion involving

1) The motor nucleus or the intrapontine course of the motor fibers (tumors, hemorrhage, etc)

2) The nerve as it crosses the posterior fossa of the skull to reach the internal auditory meatus, as in fractures of the skull or tumors in this location, the sensory portion and the auditory nerve are commonly involved as well, with loss of the sensation of taste over the anterior two-thirds of the tongue, and deafness on the affected side

3) The nerve in its course through the temporal bone as in fracture of the skull or in otitis media, inflammation of the nerve within the facial canal (aqueduct of Fallopius)—*Bell's palsy*—may occur

4) The nerve after its emergence from the stylomastoid foramen, as it lies behind the angle of the jaw, inflammation, parotid tumors, or accidental injuries may implicate the nerve in this situation

VESTIBULOCOCHLEAR NERVE

The 8th cranial nerve is a true nerve, different from the olfactory and optic nerves even though in function the three nerves seem comparable because they innervate organs of special sensation One of the major parts of this nerve contains sensory fibers from neurons located in the vestibular ganglion, with peripheral processes innervating semicircular canals, utricle, and saccule The other major part contains sensory fibers from cells of the spiral ganglion of the cochlea, their short, peripheral processes supply the hair cells of the organ of Corti A third and minor part of the nerve consists of efferent fibers to the organ of Corti, where they apparently modulate the sensitivity of the hair cells (Galambos, 1956) This group of fibers is known as the *efferent cochlear bundle*

More complete information about the functions of this nerve is given in the section on "Sensory Systems," section 8, chapter 2

GLOSSOPHARYNGEAL NERVE

The glossopharyngeal nerve contains motor, secretory, vasodilator, and sensory fibers The nerve emerges from the side of the upper part of the medulla oblongata in the groove between the olive and the restiform body Its special visceral motor fibers arise from the upper part of the *nucleus ambiguus* situated in the reticular formation of the medulla oblongata They are distributed almost entirely to the stylopharyngeus muscle, a few terminate in the circular and longitudinal muscle of the upper part of the pharynx However, division of the glossopharyngeal nerve, which is done for glossopharyngeal neuralgia, produces no motor disability so far as the patient is concerned Secretory and vasodilator fibers are the axons of cells lying in the *inferior salivatory* nucleus, which lie below the superior nucleus of the same name Via the tympanic branch of the nerve they supply the tympanic plexus, the small superficial petrosal nerve, the otic ganglion, and thence the parotid gland The sensory fibers arise from cells in the superior (jugular) and inferior (petrous) ganglia of the glossopharyngeal nerve The peripheral processes of the cells of the latter ganglion supply the taste buds of the posterior third of the tongue Those from cells of the *superior* ganglion mediate *ordinary sensations* (touch, thermal, etc), from this part of the tongue and the mucosa of the pharynx and posterior part of the mouth, the central processes (superior ganglion) terminate in the *dorsal nucleus* of the vagus and (inferior ganglion) in the lower part of the *tractus solitarius* Fibers (secondary neurons) pass from the latter nucleus and, crossing to the medial fillet of the opposite side, ascend to the thalamus, from where axons of tertiary neurons pass to the cortical area for taste

The carotid sinus nerve is another important afferent branch of the glossopharyngeal nerve It is concerned with the conduction of impulses from the carotid body and the carotid sinus to the medulla oblongata These impulses and their frequency are determined by chemical changes in the blood which are detected by the carotid body and by changes in the blood pressure which are detected by the carotid sinus These afferent impulses are concerned in reflex activities which take part in controlling the pulse rate, the respiratory rate, and the blood pressure

VAGUS NERVE

The vagus nerve contains motor, secretory, vasodilator, and sensory fibers The secretory and vasodilator fibers and the fibers to the involuntary muscle of the bronchi, heart, esophagus, stomach, small intestine, gallblad-

der, etc (parasympathetic fibers) arise from cells in the *dorsal nucleus of the vagus* (principal autonomic nucleus) This gray mass is the principal autonomic motor mechanism in the brainstem It belongs to the *general* visceral efferent group of nuclei, and extends upwards from the lower closed part of the medulla oblongata to beneath the floor of the 4th ventricle at the level of the striae medullares The *special* visceral motor fibers arise in close relationship with the motor fibers of the glossopharyngeal nerve, namely, from the cells of the *nucleus ambiguus* lying below the glossopharyngeal neurons They supply (through the superior laryngeal branch) the cricothyroid and arytenoid muscles of the larynx, and the inferior constrictor of the pharynx The pharyngeal and recurrent laryngeal branches of the vagus also convey motor fibers many of which are derived from the bulbar nucleus of the accessory nerve (see below) to the pharyngeal muscles (with the exception of the stylopharyngeus), and to the muscles of the soft palate (except the tensor palati) and larynx (except the cricothyroid) The cell bodies of the *sensory fibers* lie in the inferior ganglion of the vagus (ganglion nodosum) The peripheral processes of these cells convey impulses from the lungs, heart, larynx, pharynx, esophagus, stomach, small intestine, and gallbladder They also, through the anterior laryngeal branch, innervate the taste buds of the epiglottis and valleculae (the depressions lying at the sides of the fold running from the epiglottis to the base of the tongue) The taste fibers end centrally by synapsing with cells in the *gustatory nucleus* lying in the upper and medial part of the *tractus solitarius* These impulses are relayed upwards along the same paths as those conveying other taste impulses Afferent vagal fibers from visceral structures terminate in the dorsal nucleus This latter is, therefore, both motor and sensory in function and constitutes an important visceral reflex center It contains the cardioinhibitory and vomiting centers

Vagal afferent filaments travelling in Arnold's nerve mediate the general sensations of the skin lining the external auditory meatus and a small area behind the auricle Irritation of these fibers may cause reflex coughing These sensory fibers have their cell stations in the jugular ganglion

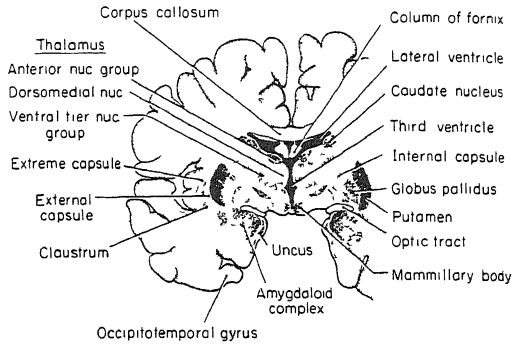


Fig 9 33 Cross section of cerebrum through thalamus and basal ganglia to show relations of these structures corpus callosum ventricular system and other portions of the brain Mammillary bodies are in caudal part of hypothalamus From Truex and Carpenter 1969

SPINAL ACCESSORY NERVE

The spinal accessory nerve is entirely motor and is made up of a bulbar and a spinal root The *bulbar* or special visceral motor root arises from the lower (caudal) end of the nucleus ambiguus from cells situated below those which give origin to the motor fibers of the vagus The bulbar fibers join the vagus within and below the jugular foramen and are distributed, as already mentioned, in the pharyngeal and recurrent branches of the latter nerve, these fibers of the spinal accessory nerve innervate the muscles of the larynx, with the exception of the cricothyroid, the muscles of the pharynx and those of the soft palate, with the exception of the tensor palati (which is supplied by the 5th cranial nerve) The *spinal* root is somatic motor and is composed of the axons of a group of cells in the anterior gray column of the spinal cord extending from the 1st to the 4th or 5th cervical segment inclusive These fibers supply the sternocleidomastoid and trapezius muscles The spinal part of the nerve exchanges fibers with the bulbar part in the jugular foramen

HYPOGLOSSAL NERVE

The hypoglossal nerve is also said to be purely motor, although it may carry proprioceptive fibers from the tongue. Its fibers are derived from a somatic motor nucleus situated near the midline in the floor of the posterior part of the 4th ventricle and medial to the nucleus ambiguus It supplies the thyrohyoid, styloglossus, hyoglossus and genioglossus muscles, and the intrinsic muscles of the tongue

Thalamus

This large gray mass is related medially to the 3rd ventricle which lies between the thalami of the two sides The thalami are joined across the midline by an isthmus, the *massa intermedia* (Fig 9 20) The posterior limb of the internal capsule lies upon the outer side of the thalamus and separates it from the lentiform nucleus Above the thalamus is the lateral ventricle, a part of whose floor it forms (Fig 9 33) In front is the head of the caudate nucleus, the arched *body* of the latter is related to the upper part of the lateral surface of the thalamus Below the thalamus are the corpus of Luys (subthalamic nucleus) and the forepart of the red nucleus

NUCLEAR GROUPS

There are five main nuclear masses in the thalamus

A) Nuclei of the Medial Mass. *1) Anterior Nuclei* These form a mass which bulges into the lateral ventricle (Fig 9 34) They receive fibers from the mammillary bodies (tract of Vicq d'Azyr), and they send fibers to the paracentral lobule and the posterior part of the cingular gyrus on the medial aspect of hemisphere *2) Dorsomedial Nuclei* These consist of a dorsolateral group of small cells, and a medial collection of large cells The former project to prefrontal areas of the cerebral cortex The large-celled portion is connected by both afferent and efferent fibers with the hypothalamus, it also projects to the corpus striatum, and has rich afferent connections with other thalamic nuclei The dorsomedial nuclei are said to serve as an association center for visceral and crude somatic sensations

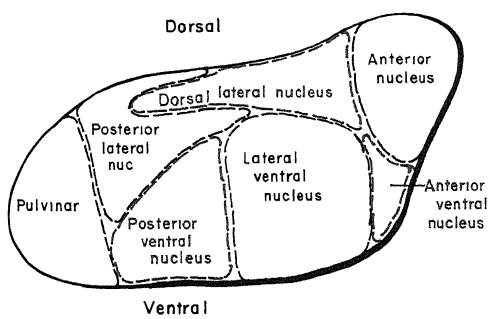


Fig 9 34 Drawing that shows relations of main divisions of the thalamic nuclei

B) Nuclei of the Lateral Mass *1) Ventral Group* This consists of *a) Anterior ventral nucleus* (also known as the ventral anterior nucleus) which receives fibers from the globus pallidus and projects to different parts of the corpus striatum, but not to the cerebral cortex This nucleus occupies the most anterior (rostral) extremity of the lateral mass *b) Lateral ventral nucleus* (the ventral lateral nucleus) The fibers entering the anterolateral part of this nucleus arises in the globus pallidus, whereas those ending in the posterior and medial part of this nucleus are derived mainly from the dentate nucleus of the contralateral half of the cerebellum (via the dentatothalamic and the dentatorubrothalamic tracts) Its efferent fibers pass to area 4, and in much smaller numbers to area 6 of the precentral motor cortex Thus, a pathway is established through which voluntary movements can be brought under the influence of the basal ganglia and the cerebellum *c)* The *posterior ventral nucleus* (ventral posterior nucleus) is the main subcortical center for sensory impulses ascending in the trigeminal, medial and spinal lemnisci (proprioceptive from muscles and joints, light touch, discrimination of two points, pain, heat, and cold) It projects to the postcentral gyrus (somesthetic area), areas 1, 3, 2, 5, and 7 This nucleus also sends fibers to the hypothalamus and corpus striatum The posterior ventral nucleus is composed of two subsidiary parts *i)* the *posteromedial nucleus*, which receives the trigeminal fibers, and *ii)* the *posterolateral nucleus* which is the thalamic station for the medial and spinal lemnisci Within this nucleus there is a definite somatotopic organization, the leg is represented most laterally, the face most medially, and the trunk and upper extremity in between

2) Dorsal Group This is also called the *lateral nucleus of the thalamus* It is subdivided into an anterior and a posterior portion, designated, respectively, the *dorsolateral* and *posterolateral nuclei*

C) The Nuclei of the Midline are, as mentioned above, situated in the massa intermedia, and in the adjacent part of the medial mass forming the upper part of the wall of the 3rd ventricle They are phylogenetically the oldest of the thalamic nuclei and are a center for the most primitive forms of sensation, e g , from the viscera and other structures oc-

cupying the axial regions of the body The midline nuclei are connected by many fine myelinated fibers with other subcortical centers, namely, the hypothalamus and midbrain nuclei They also receive fibers from the corpus striatum These nuclei have many intrathalamic connections but few if any fibers pass to the cerebral cortex, corpus of Luys (subthalamic nucleus), and the forepart of the red nucleus

D) The Intralaminar Nuclei are scattered groups of cells within the internal medullary lamina The connections of these nuclei are imperfectly known, but they are thought to receive both trigeminal fibers and fibers of the medial lemniscus, they are believed to belong to the reticular activating system They project to the globus pallidus and have many connections with neighboring thalamic nuclei A well-defined nucleus in the middle of the thalamus, and usually classed with this group is known as the *central* or *centromedian nucleus* It is especially well marked in primates, and is possibly an intrathalamic integrating center

E) The Pulvinar is the expanded continuation posteriorly of the lateral nuclear mass It overhangs the superior colliculus It projects to the cortex, its inferior part to an area comprising part of the temporal lobe, and the anterior part of the occipital lobe The rest of the pulvinar projects mainly to the parietal and temporal lobes

FUNCTIONS OF THE THALMUS

Known functions of the thalamus can be summarized under three headings, listed here in order of the amount of information available about each one

1) Sensation The thalamus and metathalamus contain the highest subcortical stages of projection of all sensory pathways of the body, with the exception of those serving olfaction Pathways from the body and head project via medial, spinal, and trigeminal lemnisci to the posterior ventral nucleus, whereas those from the eyes project to the lateral, and those for audition to the medial, geniculate body The principal pathways to the cortex, in turn, arise from these same diencephalic mechanisms A fuller account is given in section 8

2) Arousal Nuclei of the midline and the intralaminar nuclei are believed to constitute an essential part of the ascending activating system or the arousal mechanism, that originates in the reticular formation of the tegmentum and projects to cerebral cortex to keep it in a waking stage The cortical projections of these nuclei, however, are not obvious This function is discussed also in connection with the phenomenon of sleep, p 9-151

3) Association or *modulation* Practically all thalamic nuclei receive fibers from certain related parts of the cerebral cortex The discovery of these pathways that run from the cortex into the thalamus, in addition to the thalamic-cortical projections of the main sensory paths, has suggested that the cortex and thalamus somehow function together rather than independently as separable levels or stages in control or sensory systems When the word, "association," is used to describe this function, it does not have a technical meaning, but is used to describe what is not clearly motor or sensory In spite of the mysterious nature of these interrelations, the pathways are there, and they may be taken as evidence that there are, indeed, thalamic-cortical-thalamic interactions, or possibly those that are corticothalamic-cortical, and that perhaps modulate the activity of both the diencephalic and telencephalic levels of function This possibility is mentioned again later with reference to functions of the cerebral cortex, p 9-118

Hypothalamus

The hypothalamus is the anterior or ventral part of the diencephalon It is composed of neurons and their processes, both myelinated and unmyelinated, that lie in the walls and floor of the 3rd ventricle, and that form the infundibulum and pars nervosa of the pituitary gland It is bounded laterally by a plane marking the medial tips of the internal capsule, superiorly its boundary is the so-called hypothalamic sulcus in the wall of the 3rd ventricle Rostrally it ends at the plane joining the anterior commissure and optic chiasma, but histologically there is no difference between the tissue that lies above this boundary and belongs to the telencephalon, and that below the boundary in the diencephalon Consequently, the preoptic region of the telencephalon is often treated as a part of hypothalamic mechanisms In a similar fashion the caudal boundary of the hypothalamus is not sharply demarcated, it passes gradually into the tegmentum of the mesencephalon As noted ear-

her, the bulbar facilitatory mechanism extends through this region of transition and into the lateral hypothalamus

LEVELS IDENTIFIED BY EXTERNAL LANDMARKS

On the inferior surface of the hypothalamus there is a series of characteristic structures that serve as landmarks for identification of sections through this part of the brain (Fig 9 20) as follows

1) Rostrally is located the *optic chiasma*, the region over it is known as the suprachiasmatic level of the hypothalamus

2) Behind the chiasma is a prominence called the *tuber cinereum* ("grey swelling") created by the presence of the ventromedial nuclei near the base of the hypothalamus Between them arises the pituitary stalk (infundibulum) containing fibers joining nuclei of the hypothalamus and the posterior lobe (pars nervosa) of the pituitary gland The inferior surface of the stalk where it joins the hypothalamus proper, and where the vestigial cavity of the stalk opens into the 3rd ventricle, is known as the *median eminence*, it lies in the floor of the 3rd ventricle This part of the hypothalamus is known as the tuberal level

3) Behind the tuber on each side there is a prominent collection of nuclei known as the mammillary bodies because of their external form Above them the portion of the hypothalamus that is continuous through all of its levels undergoes its unmarked transition to become the mesencephalic tegmentum At this level the hypothalamus, proper, lies dorsad to the mammillary bodies much as the mesencephalic tegmentum lies above the bases pedunculorum

MICROSCOPIC ANATOMY

The hypothalamus also can be divided longitudinally into a medial portion containing neurons and unmyelinated fibers, continuous caudally with the central grey substance of the mesencephalon, and a lateral portion containing cells and both myelinated and unmyelinated fibers, continuous caudally with the reticular substance of the brainstem Rostrally the lateral region passes without obvious change into a subcortical reticular substance of the orbital portion of the frontal lobes

Two of the principal relations of the hypothalamus are with the pituitary gland (Fig 9 35) First, unmyelinated fibers from the supraoptic and paraventricular nuclei, as well as from other neurons, are gathered through the median eminence into the infundibulum and thence run to terminations in the posterior lobe (Some end in median eminence and the stalk) These are neurosecretory fibers, releasing antidiuretic and oxytocic hormones into the blood perfusing the posterior lobe (sect 7, chap 3) Second, other neurosecretory fibers not specifically identified as yet end in the ventral portions of the hypothalamus including the median eminence, where hormones known as "releasing factors" are secreted into capillaries perfusing this portion of the brain The blood from these capillaries is gathered into a second set of vessels, the hypophysial portal system, that runs along the stalk to the anterior lobe (pars distalis) of the gland The presence of these releasing factors in the blood perfusing the pars distalis controls the secretory functions of this important endocrine organ (sec 7, chap 2)

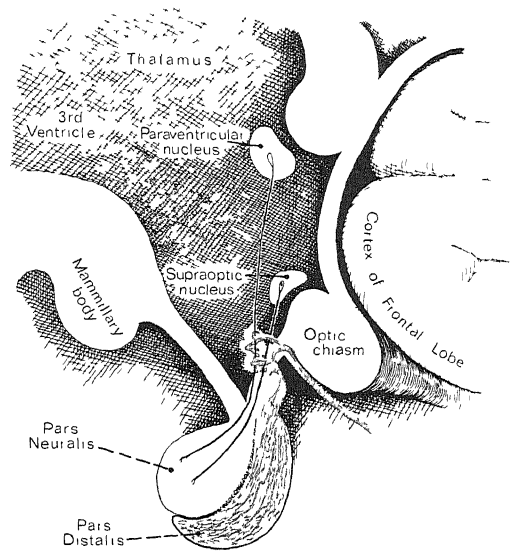


Fig 9 35 Drawing to show neural connections between hypothalamus and pars neuralis with vascular link between ventral hypothalamus and pars distalis the hypophysial portal system

Nuclei. Cells of the preoptic region are sometimes called the *preoptic* nucleus, they lie in the region of junction of telencephalon and diencephalon, and their level of origin is uncertain More caudally are found the large, deeply staining cells of the *paraventricular* and *supraoptic* nuclei The supraoptic nucleus is a

column of cells that lies over the lateral part of the chiasm and then extends back over the hypothalamic portion of the optic tract The paraventricular nucleus has the shape of a disc lying in the wall of the 3rd ventricle These nuclei are conspicuous in any stained cross-section through the rostral hypothalamus They have a rich blood supply, and they send their processes into the neural lobe by way of the median eminence and stalk, as just described In the tuberal region of the hypothalamus are two identified cell groups, the *dorsomedial* nucleus and the *ventromedial* nucleus Connections and function of both are obscure, it is known, however, that lesions in or inferior lateral to the ventromedial nuclei lead to hyperphagia and obesity In the caudal portion of the hypothalamus above the *mammillary* nuclei is the *posterior* nucleus on each side, its connections are uncertain The mammillary nuclei are of unknown function, in spite of the fact that they can be observed so easily and their connections are relatively well known (see next paragraph)

Fiber Connections. The hypothalamus receives fibers from the globus pallidus (pallidohypothalamic tract), from the amygdaloid nucleus, and through the *medial forebrain bundle*, from the olfactory lobe and the parolfactory area The medial forebrain bundle sweeps through the hypothalamus and in its course gives off fibers to several of the hypothalamic nuclei The hypothalamus, especially the supraoptic and paraventricular nuclei, also receives fibers from prefrontal areas and the precentral motor area (6) of the cerebral cortex (chap 6) both directly and indirectly through the thalamus There are probably afferent fibers from the spinal cord, doubtless similar to the direct and collateral ascending afferents which go to the bulbar part of the reticular formation It also projects through the anterior and the dorsomedial nuclei of the thalamus to prefrontal areas The most prominent tract from the hypothalamus to the thalamus is the mammillo-thalamic tract (of Vicq d'Azyr) from the medial mammillary nucleus to the anterior nuclei of the thalamus Other efferent tracts pass backwards in close relation to the ependyma of the 3rd ventricle, and descend in the gray substance of the midbrain (surrounding the cerebral aqueduct), pons, medulla (beneath the floor of the 4th ventricle), and spinal cord

The various hypothalamic nuclei are in communication with one another through fiber tracts, the best known of these is the paraventricular-supraoptic tract, but others undoubtedly exist

General Structure and Divisions

The cerebellum ("small brain") is the dorsal portion of the metencephalon, and lies in the posterior fossa of the cranium superior to the tegmentum of the mesencephalon, metencephalon, and medulla oblongata It is connected with the brainstem by the superior, middle, and inferior peduncles (Fig 9 36) at these same three levels, respectively, although the fact that the three peduncles seem to be gathered as a stalk gives the impression that its connections are mainly with the metencephalon lateral to and above the pontile tegmentum The cerebellar surface has a characteristic patterning, being divided by parallel and curved furrows into numerous laminae or folia ("leaves") Within each folium the arrangement of neurons of the cerebellar cortex is a significant feature of cerebellar organization, and will be described in some detail later The total area of the cerebellar cortex as it is folded in and out to form the folia is said to be about 100,000 mm ^{2}, a figure that is less than half the area of the cerebral cortex

When it is examined externally the cerebellum seems to have subdivisions, which have been given anatomical names (Fig 9 37) The physiological significance of many of these di-

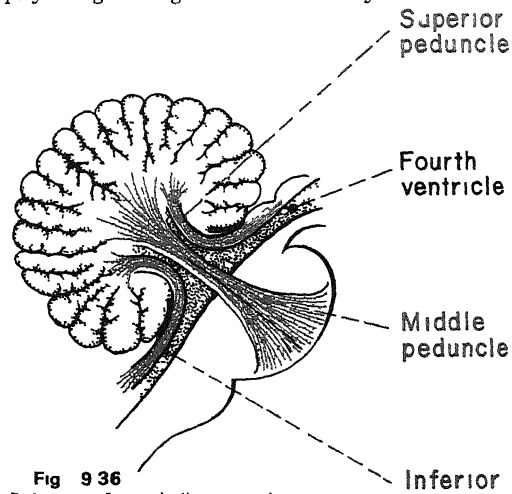


Fig 9 36
Relation of cerebellum to the brainstem and to its three peduncles

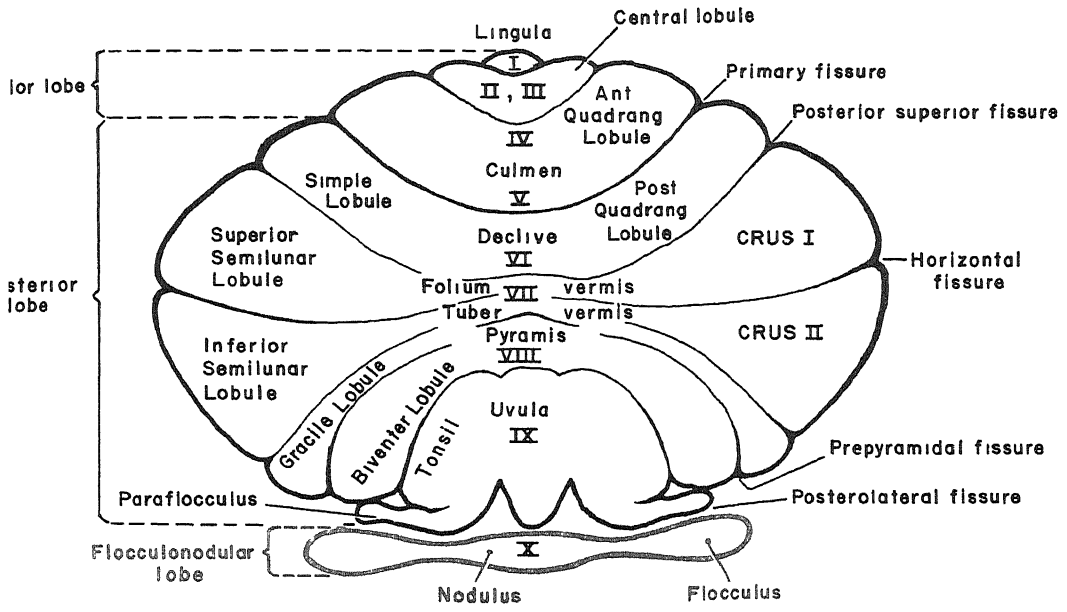


Fig 9 37 Subdivisions of the cerebellar cortex represented on an outline drawing From Trues and Carpenter 1969

ıns ıs not clear, however, and only a few of ı need to be remembered as landmarks narrow, central part of the cortex ıs called *vermis* ("worm"), and the two larger lat- masses are the right and left cerebellar *ıspheres* On the upper surface the demar- on between vermıs and hemıspheres ıs ıt Inferıorly ıt ıs more conspıcuous, and vermıs forms the roof of a depressıon wn as the *vallecula* that fits over the me- a oblongata Also ınferıorly ıs found an gated, somewhat lobulated structure on ı side called the *flocculus* ("tuft") that ıs ınuous with the most caudal or ınferıor of the vermıs, the *nodule* Together they ı the *flocculonodular lobe*, known as the ıcerebellum and believed to be the most ent part of the cerebellum from a phyloge- c poınt of vıew Its prıncipal functıonal tıonships are wıth the vestıbular nuclei of brainstem

he posterolateral fissure separates thıs lobe ı the rest of the cerebellum, the *corpus 'belli*, lyıng anterıor and superıor to the fis- ' Thıs fissure ıs one of a serıes of land- ks that run from sıde to sıde, from hemi- ere across vermıs to the other hemısphere, that consequently dıvide the cerebellum ı regıons that are eıther rostral, ıntermedi- or caudal ın theır locatıon Thıs type of dıvısıon ıs believed to be more sıgnıficant than gross anatomical dıvısıons, both phyloge- netically, embryologically, and from the vıew- poınt of comparatıve physıology (Fıg 9 37) The most ımportant fissure (asıde from the posterolateral fissure) ıs the *fissura prıma* of Elliot-Smith that dıvides the corpus cerebellı ınto a small anterıor and a larger posterıor lobe

Anterıor Lobe Thıs portıon, known as the paleocerebellum, consists of the lıngula and two other subdıvısıons of the vermıs, and the adjacent folıa of the rostral part of each hemı- sphere

Posterıor Lobe Thıs name ıs gıven to all of the rest of the cerebellar hemıspheres, ın- cludıng fıve parts of the vermıs, the majoı por- tıons of the hemıspheres themselves (the *ansi- form lobes*), the *paraflocculus*, and the *para- medıan lobule* or tonsıl, a small compact mass of cortex lyıng on eıther sıde of the ınferıor vermıs The ansıform lobule reaches ıts greatest development ın the human braın Its functıon ıs believed to be concerned wıth tonus adjustments and the muscular coordınatıon requıred for skılled movement

In hıgher anımals the posterıor lobe ıs fur- ther subdıvıded by a fissure known as the *sulcus* or fissura *prepyramıdalıs* Thıs ıs some- tımes used as a landmark for dıstınctıon be- tween the mıddle lobe of Ingvar lyıng between the fissura prıma and the sulcus prepyrami-

dalis, and a truly posterior lobe behind this sulcus and ahead of the posterolateral fissure Classified in this fashion the order of principal lobes and fissures from front to back is as follows

Anterior lobe (lingula, etc)
Fissura prima
Middle lobe of Ingvar
Sulcus prepyramidalis
Posterior lobe (pyramis, uvula)
Fissura posterolateralis
Flocculo-nodular lobe

The middle and posterior lobes, especially the ansiform lobes, belong to the neocerebellum, found only in higher animals In phylogenetic development they appear at about the same time as the cerebral cortex, the corticospinal and corticopontile tracts, and the pons The neocerebellum, cerebral cortex and pons are absent or rudimentary in forms below mammals

Neurons of the Cerebellum

When sectioned in the sagittal plane each hemisphere of the cerebellum presents a branching core of white matter which from its foliage-like appearance, has been named the *arbor vitae* The terminal branches of the white matter are covered with a coating of gray substance which constitutes the cerebellar cortex The leaflike structures so formed are spoken of as folia, and are responsible for the laminated appearance of the cerebellar surface (Fig 9 36) Unlike the cortex of the cerebrum all areas of the cerebellar cortex show a uniform histologic structure

CEREBELLAR CORTEX

The cells that lie in the cerebellar cortex have a pattern similar to the pattern of wall paper in that it is repeated and repeated throughout all of the folia (Fig 9 38) The units of each pattern include two types of fibers entering the grey matter, five types of neurons in the cortex, the glial cells, and the axons of the one cell (the Purkinje cell) that leave the cortex en route to cerebellar nuclei and brainstem This arrangement of elements was described by the great Spanish neuroanatomist, Ramón y Cajal, (1888) who studied sections stained with silver nitrate and examined under a conventional light microscope That more recent investigations using electron microscopy have not essentially altered the picture he described is a tribute to Cajal's skill and patience in preparing and examining sections

Three layers of the cortex have their characteristic appearance because they contain different proportions of cells and fibers, as follows

1) The *molecular* or plexiform layer (under a microscope it seems to have a very fine structure) is outermost It consists of unmyelinated fibers that enter the cerebellum from the inferior olive (the *climbing fibers*), axons from granule cells of the third layer, and dendrites of cells including the Purkinje cells (second layer) and Golgi type II cells (third layer) In the superficial part of the molecular layer where axons of granule cells are the most numerous there are only a few neurons, the small, star-shaped *superficial stellate* cells Deeper in this layer the stellate cells are larger, and they possess axons that run for some distance through this layer but across the folium at right angles to its long axis By means of collaterals they arborize around the bodies of a number of Purkinje cells located in a single plane of the folium (see Fig 9 38) Because of the appearance of these terminals, these stellate cells are known as *basket* cells

2) The *intermediate* layer of *Purkinje* cells The large, flask-shaped bodies of these cells, which are peculiar to the cerebellum, form a layer between the molecular and granular layers Terminals of the basket cells surround their cell bodies Their denrites pass outwards through the entire thickness of the molecular layer where they arborize luxuriantly, but only in two dimensions, like a fan in a single geometrical plane This plane is at right angles to the long axis of the folium The afferent *climbing fibers* make connections on these dendrites, and many of the axons of the granule cells of the innermost layer also end there The axons of the Purkinje cells enter the white substance, and end by synapsing with cells in the cerebellar nuclei and brainstem

The Purkinje cells have been referred to as the "final common pathway" of the cerebellar cortex, because it is through them that the cerebellum exerts its influence upon the central nuclei of the cerebellum, as well as upon other parts of the central nervous system As noted later, the Purkinje cells seem to be exclusively inhibitory in their action upon other neurons

3) The *granular* layer contains numerous

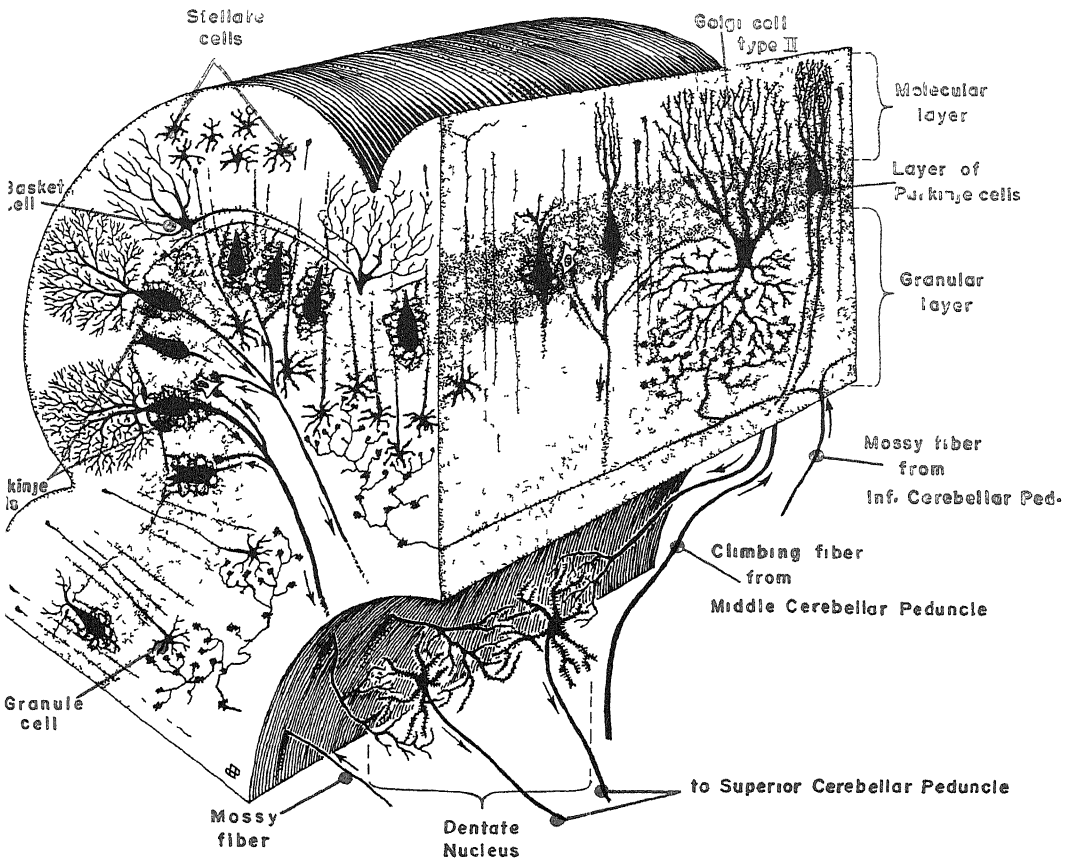


Fig 9 38 Cerebellar cortex drawn in three dimensions to show geometry of location of Purkinje cells Golgi s granule basket and stellate cells with terminals of climbing and mossy fibers

rve fibers, and the two remaining types of urons One is the *granule* cell, small, round, sely packed, and possessing four or five ndrites that end in a tuft of branches close the cell body and connect with terminals of ssy *fibers* and with axons of Golgi type II ls One long process (axon) extends out into molecular layer where it divides and sends branch in each direction along the length of folium These branches, therefore, pass ough the plane of arborization of the den- tes of many Purkinje cells The last type of uron is known as the *Golgi type II* cell be- use it has no long axon These are large cells ated in a plane just inside that of the Pur- nje cells They have dendrites that branch tensively in all planes of the molecular layer, d send their axons to branch repeatedly and d upon dendrites of the granule cells

The three types of fibers most essential to understanding of cerebellar function are the lowing

1) Mossy fibers are coarse, myelinated fibers that have a characteristic appearance as they terminate in the region of the terminals of granule and Golgi II cells in the granular layer of the cortex They are known to enter the cerebellum from the spinocerebellar systems, and are believed to come also from the pontocerebellar, vestibulocerebellar, reticulocerebellar, and other systems from the brainstem to cerebellar cortex (Eccles, et al , 1967)

2) Climbing fibers are also myelinated, but smaller, and have the unusual, ivy-like association with Purkinje dendrites already described They are believed to come largely if not exclusively from the olivary nuclei (Eccles, et al , 1967) These two types of fibers constitute the only known input to the cerebellum

3) Axons of the *Purkinje* cells are the sole output from the cerebellar cortex, they are small, myelinated fibers that run to the nuclei of the cerebellum, and to Deiter's nucleus of the vestibular complex These fibers, how-

ever, because most of them end within the cerebellum on its subcortical nuclei, are only a small fraction of the efferent output of the cerebellum as a whole Most of this output is carried by myelinated fibers originating in the central nuclei and carried to the brainstem and thalamus by superior and inferior peduncles

THE CEREBELLAR NUCLEI

The cerebellum contains on either side four separate gray masses (Fig 9 39) These are *a) nucleus fastigii* (nucleus of the roof), *b) nucleus globosus, c) nucleus emboliformis,* and *d) nucleus dentatus* (The nuclei globosus and emboliformis of primates are comparable to the nucleus interpositus of lower mammals) The first three of these are phylogenetically older than the nucleus dentatus, which is found only in mammals The fastigial nuclei, which are the most ancient of all, lie near the midline on either side in the roof of the 4th ventricle They receive fibers from the paleocerebellum (anterior lobe, pyramid, uvula, and flocculonodular lobe), and from the vestibular nuclei and eighth nerve through the inferior peduncle, they also project to the vestibular nuclei The globose and emboliform nuclei, which are placed more laterally than the roof nuclei, receive fibers chiefly from the anterior lobe The globose nucleus also receives fibers from the flocculonodular lobe Both nuclei project through the superior peduncles to the large-celled portion of the red nucleus The more lately acquired dentate nucleus, which is placed most laterally, is a large, crenated mass of gray matter, bent acutely upon itself It receives fibers from the Purkinje cells of the neocerebellum It projects through the superior cerebellar peduncle to the small-celled portion of the red nucleus (and to some extent also to the large-celled part), and to the ventrolateral group of the thalamic nuclei (Fig 9 39)

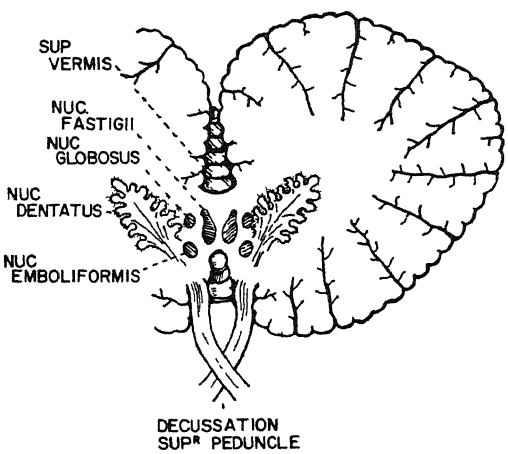


Fig 9 39 Diagram of a horizontal section through the cerebellum to show relation of cortex and subcortical nuclei, with superior peduncle and its decussation in the mesencephalon.

Connections of the Cerebellum

The white matter of the hemispheres is composed of *a) projection fibers,* i e , fibers which leave or enter the cerebellum via the peduncles, *b) association fibers,* which connect different regions of the same hemisphere, and *c) commissural fibers* connecting cortical areas of the two hemispheres

The *flocculonodular lobe,* mainly the nodule, is connected by both afferent and efferent fibers (some of which are relayed in the fastigial and globose nuclei) with the vestibular nuclei and the vestibular nerve The anterior lobes and the ansiform lobules receive spinocerebellar and pontocerebellar fibers, respectively, and project to brainstem and thalamic nuclei

The fibers, afferent and efferent, connecting the cerebellum with other portions of the nervous system are all carried in three large bundles, the *cerebellar penduncles,* having the following constitution

A) Inferior Peduncles (Restiform Bodies) they are predominantly afferent
- *1) Afferent Fibers*
  - *a)* Dorsal (posterior) spinocerebellar tract (the direct cerebellar tract)
  - *b)* Dorsal external arcuate fibers
  - *c)* Ventral external arcuate fibers
  - *d)* Vestibulocerebellar fibers
  - *e)* Olivocerebellar tract, the fibers of which end in all parts of the cerebellar cortex as climbing fibers upon dendrites of Purkinje cells
  - *f)* Fibers from the fifth cranial nerve, and possibly from the ninth and tenth, to pyramid, uvula and paraflocculus
- *2) Efferent Fibers*
  - *a)* Fastigiobulbar (or cerebellovestibular) tract
  - *b)* Cerebello-olivary tract to inferior olives of both sides

They also have an excitatory action on the basket cells and the dendrites of the Golgi type II cells According to these authors this is a complete account of excitatory effects within the cortex, they involve three kinds of axons, climbing fibers, mossy fibers, and granule cell fibers All other actions are of an inhibitory nature

The *inhibitions* they have identified include the following

1) Purkinje cells are inhibitory upon cells of subcortical cerebellar nuclei and upon Deiter's nucleus,

2) Basket cells are inhibitory upon Purkinje cells,

3) Golgi type II neurons are inhibitory upon the mossy fiber-granule cell synapse

These observations suggest that the overall influence of the cerebellar cortex upon subcortical nuclei is one of inhibition, but that this inhibition is graded by the neuronal actions that simultaneously excite and inhibit the discharge of Purkinje cells

In addition to this complex interrelation of excitation and inhibition, another interesting feature of cerebellar cortical activity is the geometric pattern it assumes, in which the orientation of the plane of arborization of Purkinie cells in the folium is like slices of bread in a long loaf The distribution of axons of basket cells is in similar planes across the long axis of the folium The distribution of parallel fibers from granule cells, however, is along the length of the folium and therefore across the planes of Purkinje cells and basket cell axons The dendrites of the Golgi type II cells branch out in all planes of the molecular layer

EXPERIMENTAL STUDY OF THE CEREBELLUM

The cerebellum seems to be relatively simple because *a*) the pattern of cortical organization is straightforward and repeated throughout the folia, *b*) the cortex receives only the input of mossy and climbing fibers, *c*) there are only five types of neurons in the cortex, and *d*) the cortex and subcortical nuclei have a constant relationship by way of axons of Purkinje cells Nevertheless, the multitude of connections between cerebellum and the spinal cord, the tegmentum of the brainstem, olivary complex, vestibular nuclei, red nucleus, thalamus, and cerebral cortex has made analysis of its functions exceedingly difficult In the first place, there is no lack of information about these possible extrinsic connections, the course of almost innumerable fibers has been described with great care, with the result that on this basis alone it is impossible to conceive of any simple pattern of cerebellar control of motor activity Second, there is no lack of observations as to the deficiencies that follow cerebellar injury in both human patients and laboratory animals Third, unlike other portions of the brain, stimulation of the cerebellar cortex in anesthetized animals does not induce motor responses that lend themselves to easy interpretations In spite of these problems, however, one should try to understand the nature of the principal cerebellar functional relations, the general character of deficits that follow ablation or injury, and possible explanations for the results of stimulation experiments This information is summarized in chapter 5 beginning on p 9-98

SUBCORTICAL MECHANISMS OF THE TELENCEPHALON

The Basal Ganglia

The basal ganglia form a large and prominent part of the human brain, and are a very important part of the nervous mechanisms controlling muscular activity (Fig 9 41) Destructive lesions in the human basal ganglia are associated with several well known, common, severe motor disorders which are fairly constant in their clinical pattern Yet, our knowledge of the nervous connections and of the functional activity of the human basal ganglia is grossly deficient They have grown proportionately from the lower primates to

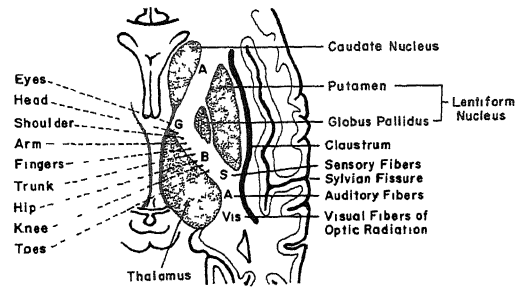


Fig 9 41 Horizontal section through right cerebral hemisphere showing relationships of thalamus internal capsule and basal ganglia including caudate nucleus, lentiform nucleus and claustrum Localization of fibers in internal capsule is also shown

man, so that whereas the globus pallidus, putamen and caudate nucleus occupy 980 cu mm in the rhesus monkey, they measure 2775 mm ^{3} in the chimpanzee and have a volume of 14,912 mm ^{3} in man (Harman and Carpenter, 1950)

NUCLEAR GROUPS

In this discussion the term "basal ganglia" will be used to include the *caudate nucleus*, the *putamen*, and the *globus pallidus* of the telencephalon, the *subthalamic body of Luys* of the diencephalon, and the *red nucleus* and the *substantia nigra* of the mesencephalon (The caudate nucleus and putamen are often referred to as the *striatum* and the globus pallidus as the *pallidum* The putamen and globus pallidus are sometimes referred to as the *lenticular* or *lentiform nucleus*) The *claustrum* should also be included in the basal ganglia but as practically nothing is known of its connections or functions it need not be considered further It might be argued that the subthalamic body, the red nucleus, and the substantia nigra should not be included, but their anatomical relationships, their numerous connections, and their apparent close functional relationships with the other nuclei of the basal ganglia make their inclusion natural and reasonable The amygdala is another subcortical collection of neurons that might be included But it has no known connections with the functions of the skeletal muscular system and seems more likely to be related to vegetative activities

The cytology of the various nuclei under discussion offers little assistance in understanding their functional activity For the most part the nuclei seem to be divided into two groups—the small-celled and the large-celled parts As is true of the cerebral cortex the small-celled parts appear to be the receptive divisions of the basal ganglia, whereas the large-celled parts have a predominantly motor function The caudate nucleus and the putamen constitute the receptive portion and are said to receive their afferent innervation from the precentral motor cortex (areas 4, 4s, and 6), the suppressor strips of Dusser de Barenne and his co-workers (areas 4s, 8, 2, and 24), (Fig. 9.42) and in a small way from the thalamus. In turn the caudate nucleus and putamen send their efferent fibers mainly to the large-celled or "motor" portion of the lenticular nucleus, the globus pallidus (Fig 9 42), with a much smaller group of efferent fibers going to the substantia nigra Similarly the substantia nigra is divided into a large-celled (pigmented) dorsal portion and a small-celled (unpigmented) ventral portion which lies just above the cerebral peduncle The red nucleus, too, is divided into a small-celled anterior portion and a large-celled posterior portion which gives rise to the rubrospinal tract

As noted above (p 9-29), in the dog and cat the magnocellular portion of the red nucleus and the rubrospinal tract are large and presumably important structures, whereas in the primates this portion of the red nucleus and the rubrospinal tract which arises from it are small and apparently insignificant (Stern, 1936, Massion, 1967) The parvocellular or main portion of the nucleus in man has a variety of connections, most of them related to the dentatorubrothalamic pathways that send fibers through and around the red nucleus on their way from cerebellar nuclei to the lateral ventral nuclei of the thalamus

CONNECTIONS

Within the basal ganglia almost all parts are connected with all other parts (Fig 9 42), i e, the globus pallidus sends fibers to the red nucleus, and substantia nigra and the subthalamic nucleus, the substantia nigra sends fibers to the red nucleus, the globus pallidus and the subthalamic nucleus, the subthalamic nucleus sends fibers to the globus pallidus, red nucleus and the substantia nigra Within the basal ganglia there are also recurrent systems, such

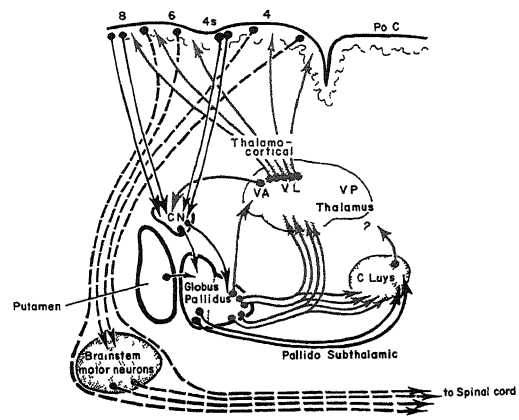


Fig 9 42 Outline of pathways by which basal ganglia are interconnected and related to motor mechanisms of cerebral cortex and brainstem and spinal cord

as the projection from the caudate nucleus to the globus pallidus, the globus pallidus to the ventral anterior nucleus of the thalamus, and from this thalamic nucleus back to the caudate nucleus (Fig 9 42) Two parts of the basal ganglia, however, remain more or less aloof from this intranuclear activity They are the caudate nucleus and putamen, and the red nucleus The caudate nucleus and putamen on the one hand receive their afferent innervation largely from the cerebral cortex (with additional afferent connections from the substantia nigra and from the ventral anterior nucleus of the thalamus), and they distribute their efferents to the globus pallidus and the substantia nigra The red nucleus receives from all of the other nuclei of the basal ganglia (and from the precentral motor cortex and the cerebellum) but sends all of its efferent fibers to structures outside of the basal ganglia, to the spinal cord, the reticular substance of the brainstem, the ventrolateral nucleus of the thalamus, the inferior olive and the oculomotor nuclei (Fig 9 43) Thus, the caudate nucleus and the putamen appear to be acting as the primary receiving station of the basal ganglia, whereas the red nucleus and the globus pallidus are acting as the principal sending stations In addition to their efferent fibers there are fibers from the substantia nigra and the subthalamic nucleus to the substantia reticularis and the centermedian nucleus of the thalamus

Internal Capsule

If one considers the very large number of cells in the cerebral cortex capable of acting upon motor mechanisms of brainstem and spinal cord, together with projections to basal ganglia, thalamus, pons, and cerebellum, it becomes obvious that pathways for transmission of these influences must be plentiful They leave the cortex via the central white substance of the hemispheres, the region that contains also the fibers ascending to the cortex, and fibers running to and from the corpus callosum As the corticofungal fibers descend toward the level of the mesencephalon, they run past the basal ganglia of the telencephalon, and also skirt the lateral and ventral sides of the thalamic collections of neurons From the level of the central white substance all the way down to the spinal cord these corticospinal pathways remain separate from the parts of the brain through which they are passing, except for collaterals given off in transit

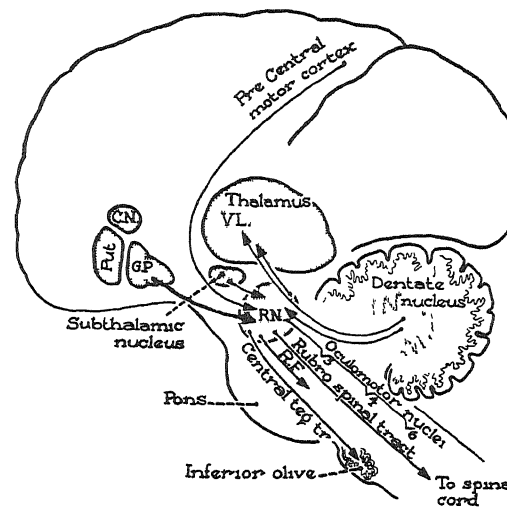


Fig 9 43 Connections of the red nucleus Abbreviations refer to principal basal ganglia *R F* denotes fibers to reticular formation

In its longitudinal organization the central nervous system contains two parallel systems One is what appears to be the modulating system that functions as a control for the other system, and that includes the reticular substance of tegmentum and hypothalamus, the thalamic nuclei of midline and intralaminar location, the basal ganglia, and the cerebral-pontile-cerebellar-thalamic complex The other includes the long ascending and descending systems, with ascending relays in dorsal grey columns of spinal cord, in the brainstem, and in the thalamus, and with descending relays that always seem to be diversions from the main tracts into nuclei or neuronal groups encountered along the way The internal capsule is the broad highway in both directions between the cerebral cortex and subcortical and spinal centers In a horizontal section of the cerebrum it is seen as a compact band of white matter lying between the thalamus and caudate nucleus on its inner aspect and the lenticular nucleus on its lateral aspect (cf Fig 9 41) In addition to the corticospinal and corticobulbar tracts the following are the principal tracts which compose it

I) Descending

a) Corticopontine tracts consisting of three bundles of fibers, *frontopontine*, *parietopontine*, and *occipitopontine* from the corresponding areas of the cerebral cortex to the nuclei of the pons

b) *Corticorubral tract*, from the frontal lobe to the red nucleus
c) *Corticothalamic* from almost all areas of the cortex to various nuclei of the thalamus
d) *Corticostriatal fibers* from the precentral area of the cortex to the caudate nucleus and the globus pallidus
e) *Occipitotectal fibers* from the para- and peristriate cortex to the tectum of the midbrain
f) *Corticonigral fibers* from the precentral and temporal areas to the substantia nigra

II) Ascending
a) *Thalamic radiations* consist of four bundles ascending from the thalamus, *anterior*, *superior*, *posterior*, and *inferior*, which connect the thalamus with various parts of the cerebral cortex The anterior group terminate in prefrontal areas, and the superior group in the somesthetic area (postcentral gyrus)
b) *Auditory radiation fibers* ascending from the medial geniculate body to the temporal cortex (geniculotemporal fibers)
c) *Optic radiation fibers* ascending from the lateral geniculate body to the visual area in the occipital cortex (geniculocalcarine) fibers
d) *Rubrocortical fibers* from the red nucleus to the cortex

The internal capsule is bent in the horizontal plane to conform a convexity directed medially The region of the angle so produced is called the knee (genu), the portions in front and behind this are known, respectively, as the anterior and posterior limbs The shorter anterior limb lies between the caudate nucleus and the lenticular nucleus, the posterior limb between the lenticular nucleus and the thalamus The extension backwards of the posterior limb is known as the *retrolenticular part*

The anterior limb is occupied by the frontopontine fibers and the anterior thalamic radiation (thalamofrontal fibers), the frontothalamic tract, corticostriatal and corticorubral tracts In the anterior three-fifths of the posterior limb are transmitted the corticobulbar and corticospinal fibers The fibers of these two tracts are organized in accordance with the portions of the body which they innervate, those carrying impulses for the eye muscles, the muscles of the tongue, and face and upper and lower extremities arranged in this order from before backwards, and those transmitting impulses destined for the proximal part of an extremity are placed in front of those for the more distal muscles The remaining part of the posterior limb, and the retrolenticular part carry in anteroposterior order the superior and posterior thalamic radiations, the auditory (geniculotemporal), and the inferior thalamic radiations, corticothalamic, and parietopontine fibers, and the optic (geniculo-occipital) radiations

The internal capsule is supplied with blood through the lateral striate branches of the middle cerebral artery, by the recurrent artery of Hubner from the anterior cerebral artery, and the anterior choriodal artery from the internal carotid artery The internal capsule is frequently a site of injury to the nervous system through hemorrhage from one of the striate branches Both sensory and motor deficits may result

3 / CONTROL OF SMOOTH MUSCLE AND EXOCRINE GLANDS

Before a description of the mechanisms controlling skeletal muscle is attempted, it will be instructive to consider the control of smooth muscle and exocrine glands, for two reasons First, these latter controls are less complex than the other, second, an explanation of the organization of the autonomic system and how it functions will include topics such as chemical transmission of nerve impulses that are relevant to all neural controls The characteristics of the autonomic system that seem to make it unique can be shown to be merely overstated displays of how neural systems are put together The one really unique feature of such systems turns out to be the fact that smooth muscle can, but skeletal muscle cannot be inhibited by a mechanism acting directly upon it

Terminology and Organization

Smooth muscle and glands are controlled by systems that have both afferent and efferent components, although in the history of studies on these systems the efferent components have been given more attention than the afferent To understand the literature on these controls, it is necessary first of all to grasp the distinctions between terms used to designate parts of the system The most comprehensive of these terms is the *autonomic* nervous system, a name utilized some seventy years ago by the Cambridge physiologist, J N Langley (1903, 1921) A few years later his colleague, W H Gaskell (1916), employed the term *involuntary* nervous system as a synonym Both of them wrote primarily of the motor components (Fig 9 44) In the most recent literature, on the other hand, both the words autonomic and also *visceral* are employed to include both sensory and motor innervations of visceral structures, specifically, of smooth muscle and glands

In the years that followed the work of Gaskell and Langley, additional names were given to the principal subdivisions of the motor side of the visceral system From this much of the confusion of terminology arose, because the ancient anatomical name, *sympathetic*, was redefined and used with *parasympathetic* to denote two supposedly opposed portions of the system These terms were popularized especially in the United States by publications from the laboratory of Walter Bradford Cannon of Harvard (1929, 1932), and thence came into general use without precise definition In response to this lack of precision, two more specific classifications were developed, one anatomical in which the two subdivisions are *craniosacral*, and *thoracolumbar*, named for the site of origin of efferent fibers The other classification was originally pharmacological, with *adrenergic* and *cholinergic* utilized to identify fibers with reference to their chemical transmitter Unhappily, these three sets of terms cannot be used interchangeably, although they have superficial characteristics that suggest that they might be The term, sympathetic, has come to have a philosophical, almost mystic meaning It is to a degree synonymous with the thoracolumbar division, and many of its terminals are adrenergic But it contains also certain important cholinergic terminations The craniosacral division is the same as the parasympathetic portion, and all of its fibers are cholinergic These usages are summarized as follows

Terminology	Divisions	
Anatomical	Craniosacral	Thoracolumbar
Physiological	Parasympathetic	Sympathetic
Pharmacological	Cholinergic	Cholinergic and Adrenergic

Autonomic control systems are constituted from detector (afferent) neurons, connector or interneurons, and efferent or effector neurons in the fashion outlined in chapter 1. There are two curiosities, however, about the location of these neurons Whereas interneurons in general are located exclusively in spinal cord or brain, the last interneuron of the autonomic motor system sends its process through the ventral roots or via a cranial nerve to termi-

nate in one of the ganglia of the autonomic system (Fig 9 45) This means that without exception the efferent or effector neuron has its location in one of these ganglia—either in a ganglion of the sympathetic chain (cervical, thoracic, lumbar, or sacral), in a so-called "collateral" ganglion lying near a major blood vessel (e g , coeliac, superior mesenteric), or in a terminal ganglion lying in, on or near the viscus innervated These relations, again, sometimes give rise to confusion, inasmuch as the connector neuron is known as the *preganglionic* neuron, the effector as the *postganglionic* neuron, and homology with the somatic motor system is thereby obscured

AFFERENT VISCERAL NERVES

Impulses are transmitted from the viscera by afferent fibers which pass through the various plexuses and reach the central nervous system via the vagus, pelvic, splanchnics, and other autonomic nerves

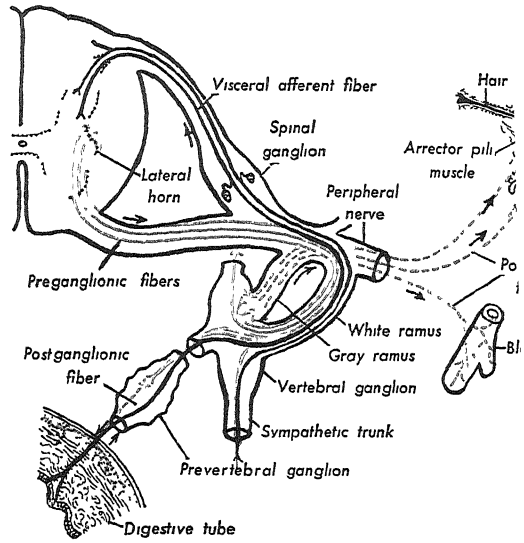


Fig 9 45 Diagram showing course of visceral afferent and efferent fibers in dorsal and ventral roots respectively and their distribution through the thoracolumbar system Copenhaver cited from Truex and Carpenter 1969

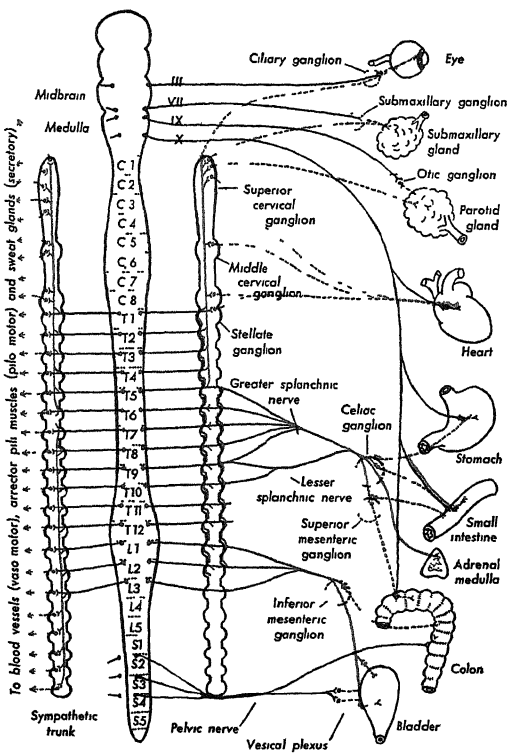


Fig 9 44 Diagram of connections of autonomic or visceromotor system Craniosacral components are in black, thoracolumbar connections are in brown Course of postganglionic fibers is shown by broken lines (Truex and Carpenter 1969)

The *afferent fibers of the thoracolumbar division* are the peripheral processes of ganglion cells in the posterior spinal root ganglia from the 1st thoracic to the 3rd lumbar segments None arises from sympathetic ganglia They reach the sympathetic trunk via the white rami communicantes and are distributed to the viscera along with the corresponding efferent fibers (Fig 9 45) Though some of these afferent fibers are nonmedullated, the majority are medullated, and of larger size than the efferent fibers The ganglion cells of the posterior roots which give origin to the visceral afferent fibers have not been shown definitely to differ from those giving rise to the ordinary somatic sensory fibers. The afferent fibers pass to their destinations without interruption in any of the sympathetic ganglia, passing directly to the viscera in the splanchnics or the other visceral nerves A certain proportion also enter the spinal nerves for distribution to the limbs via the gray rami

The *afferent fibers of the* cranial division are the peripheral processes of cells in the ganglia of nerves VII, IX, and X The central processes of these neurons terminate in the nucleus of the tractus solitarius The *afferent fibers of the pelvic nerve* arise from cells in the posterior root ganglia of the 2nd, 3rd, and 4th sacral

nerves They pass peripherally with the efferent autonomic fibers

A certain degree of independent reflex activity has been attributed to the intrinsic plexuses, e g , of the intestine, when these are separated from the central nervous system by division of the main autonomic nerves It is also true that some independent activity can be carried out through axon reflexes or possibly through some of the more peripherally placed ganglion cells Other parts of the autonomic system, however, cannot function apart from the central nervous system The larger ganglia of the parasympathetic or sympathetic, for instance, do not serve as reflex centers It is clear from the description of the origin and course of the visceral afferent nerves given above that no anatomical basis for such action exists

Sympathetic Motor Division

Cells of origin of the sympathetic division are situated in the lateral horns of the spinal cord (intermediolateral cell column) from the 8th cervical or 1st thoracic to the 2nd or 3rd lumbar segments The axons of these preganglionic cells leave the cord by the corresponding anterior nerve roots and synapse with postganglionic nerve cells in one or another of the outlying ganglia Preganglionic fibers are myelinated, postganglionic fibers have no myelin sheath

The ganglion cells are all motor or secretory in function, they have no afferent connections from the periphery Afferent fibers from the various viscera and blood vessels that pass through the sympathetic nervous system have their ganglion cells in the posterior root ganglia, as described above

The motor ganglia are arranged in three systems or groups *A) paravertebral, B) prevertebral* (or *collateral*), and *C) terminal* (or *peripheral*)

PARAVERTEBRAL GANGLIA AND THE GANGLIONATED CORD

Ganglia of the paravertebral group lie in close relation to the vertebral bodies On each side, there is a series of some 22 ganglia connected together by intervening fiber tracts to form a nodular cord extending from the base of the skull to the front of the coccyx This is known as the *sympathetic chain* or the *ganglionated cord of the sympathetic nervous system* All but the lowest four of these (the sacral ganglia) belong to the thoracicolumbar division It will be described in sections

Cervical Part of the Sympathetic Chain The cervical part of the sympathetic chain possesses three ganglia—the *superior, middle,* and *inferior cervical ganglia* They are relatively large and are believed to result from the fusion of two or more smaller ganglia

The Superior Cervical Ganglion, situated below the base of the skull, is the largest of the three It receives preganglionic fibers from the upper thoracic segments of the spinal cord, and its cells supply fibers (postganglionic) to the vessels, glands, and cutaneous muscle of the head It is probably formed by the fusion of the uppermost three or four cervical ganglia

Middle Cervical Ganglion is formed presumably by the coalescence of the 5th and 6th cervical ganglia Its branches include fibers to the *5th and 6th cervical nerves* and thence to the blood vessels, sweat glands, and cutaneous muscle within the area of distribution of these nerves, *the middle cardiac nerve* to cardiac plexuses, and branches which extend along the inferior thyroid artery to the *thyroid* and *parathyroid glands*

Inferior Cervical Ganglion probably represents the union of the 7th and 8th cervical ganglia, in the dog and cat, and in most human subjects, it is usually fused with the first thoracic ganglion, and occasionally with the second as well, to form the so-called *stellate ganglion* It gives off branches to the *7th and 8th cervical* and the *1st thoracic nerves,* sometimes also to the 6th cervical and the 2nd thoracic, the *inferior cardiac nerve* to cardiac plexuses, and branches which form plexuses upon the subclavian artery and its branches Sympathetic fibers are thus carried into the cranial cavity along the vertebral artery, and over the axillary and commencement of the brachial

Thoracic, Lumbar and Sacral Ganglia. The *thoracic ganglia* are 10 or 12 in number on each side They are evenly spaced, one to each spinal segment As mentioned the first thoracic and inferior cervical ganglia are commonly partially or completely fused to form an irregularly shaped mass, called the *stellate ganglion* There are usually 4 *lumbar,* in addition to 4 or 5 *sacral* ganglia The sacral portions of the two sympathetic trunks converge

below and fuse upon the anterior surface of the coccyx to form a terminal swelling—the *coccygeal ganglion* or *ganglion impar*

PREVERTEBRAL (OR COLLATERAL) GANGLIA

These lie in the thorax, abdomen, and pelvis in relation to the aorta and its branches (Fig 9 46) The larger of the prevertebral ganglia are *a*) the *celiac* (*solar* or *semilunar*) *ganglion*, lying in relation to the origin of the celiac artery, *b* the *superior mesenteric ganglion*, situated below the origin of the superior mesenteric artery, and *c*) the *inferior mesenteric ganglion*, which bears a corresponding relation to the inferior mesenteric artery, this ganglion is rarely present in man

TERMINAL GANGLIA

These consist of small collections of ganglion cells situated in close relation to the innervated organs, especially those of the pelvis, e g , the bladder and rectum

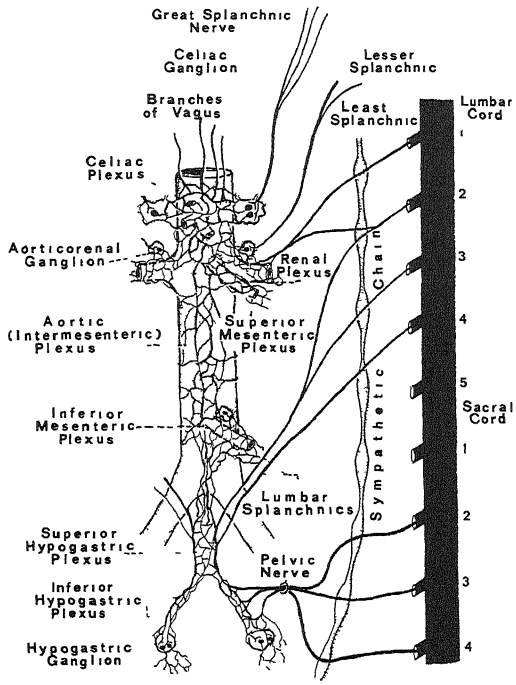


Fig 9 46 Diagram of nerve plexuses of the abdomen and pelvis

OUTFLOW OF SYMPATHETIC FIBERS FROM THE CENTRAL NERVOUS SYSTEM

It has already been stated that the cells giving rise to the thoracicolumbar, preganglionic fibers are situated in the thoracic and upper lumbar segments of the cord This limited region (8th C or 1st T to 2nd or 3rd L inclusive) constitutes the only outlet for sympathetic impulses The fibers emerge from the spinal cord through the anterior root of the spinal segment in which their cell bodies are placed In a cross-section of the anterior root they appear as fine medullated fibers (2 5μ or less in diameter) intermingled with the larger, medullated, somatic (motor) fibers They separate almost immediately, however, from the somatic motor fibers of the anterior root, and enter the corresponding ganglion of the sympathetic chain Thus, the spinal nerves from the 8th cervical or 1st thoracic to the 2nd or 3rd lumbar, but not others, are connected each to a paravertebral ganglion by a delicate white strand composed of preganglionic fibers and known as the *white ramus communicans* (plural, *rami communicantes*) A preganglionic fiber after entering the ganglion may pursue one of three courses it may *a*) form synapses with cells in the ganglion which it first enters, or *b*) traverse the ganglionated cord without interruption to find a cell station in either a prevertebral or a terminal ganglion, or *c*) pass up or down the sympathetic trunk for some distance to terminate in a ganglion at a level higher or lower than the segment from which it originated It may give off collateral branches to ganglion cells along its course In any event, the preganglionic fibers issuing from a given segment connect with several ganglia (from five to nine) Furthermore, each preganglionic fiber may form a large number of synapses within a given ganglion Ranson and Billingsley (1918) found that in the case of the superior cervical ganglion each fiber communicated with some 32 ganglion cells, this accounts for the diffuse nature of the sympathetic discharge

DISTRIBUTION OF SYMPATHETIC FIBERS TO THE PERIPHERY

Ganglia of the sympathetic chain are connected with the spinal nerves supplying the limbs and trunk by delicate filaments called *gray rami communicantes* They are composed of the axons of the ganglion cells in the para-

vertebral ganglia, and are therefore called *postganglionic* The gray rami in the thoracic and upper lumbar regions join the spinal nerves close to the points at which the white rami arise Their constituent fibers are continued in the peripheral nerves for the supply of the blood vessels, sweat glands, and smooth muscle of the skin Whereas, as already mentioned, only a limited number of spinal nerves possess white rami, *every spinal nerve receives a gray ramus* It therefore follows that sympathetic impulses going to parts supplied by upper cervical, lower lumbar or sacral somatic nerves, must travel considerable distances up or down the sympathetic trunk before reaching an outlet through a gray ramus To the limbs and trunk the preganglionic fibers to the upper limb arise from the 2nd to the 7th thoracic spinal segments, most of these fibers are contained in the white rami of the 2nd and 3rd thoracic anterior roots The postganglionic fibers to the upper limb may arise from the upper 1st to 4th thoracic, and the inferior and middle cervical ganglia, those for the lower limb from the lumbar and sacral ganglia Excision of the inferior cervical or the stellate ganglion when present and the upper three or four thoracic ganglia provides as complete a postganglionic sympathectomy as it is possible to make of the upper limb, as well as the head and neck It also will destroy most of the postganglionic fibers to the heart This operation, however, may fail to remove completely the sympathetic supply to the arm since there may be "intermediate" ganglia which are not removed in such an operation

Parasympathetic or Craniosacral Motor Division

The cells giving rise to parasympathetic fibers are situated at three different levels of the central nervous system—the *midbrain*, the *medulla oblongata*, and the *sacral region of the spinal cord* The axons of these cells leave the central nervous system to connect with ganglion cells lying within or in close relation to the innervated organ As in the case of the sympathetic division the axons of the central cells are called *preganglionic*, those of the ganglion cells, *postganglionic* The former are medullated, the latter nonmedullated The three levels from which parasympathetic fibers emerge will be referred to as the *tectal* (or *midbrain*), the *bulbar* and the *sacral* outflows, respectively

TECTAL OR MIDBRAIN OUTFLOW

The group of cells composing the Edinger-Westphal nucleus of the oculomotor nerve in the mesencephalic part of the intermediate motor column give rise to fibers that are conveyed in the third nerve as far as the ciliary ganglion where they find their cell stations Postganglionic fibers emerge from the ganglion in the short ciliary nerves, and terminate in the sphincter pupillae and the ciliary muscles

BULBAR OUTFLOW

These fibers leave the brain in the *facial*, *glossopharyngeal*, and *vagus nerves*

1) The parasympathetic fibers (secretory and vasodilator) entering the *facial nerve* arise from the *superior salivatory* (*salivary*) *nucleus*, which lies dorsal and lateral to the lower end of the motor nucleus of the facial [1] These fibers emerge from the brain in the sensory root of the facial nerve (*nervus intermedius*) and travel with the latter to the facial canal of the temporal bone Here they leave the facial *i*) They may join its *chorda tympani* branch which later joins the lingual to be conveyed to the floor of the mouth At this point some of the chorda fibers (secretory and vasodilator in function) separate from the lingual again to enter the *submaxillary* (*submandibular*) *ganglion* from where they are relayed, the postganglionic fibers pass to the submaxillary and submandibular glands, and the mucous membrane of the mouth *ii*) They may run with the great superficial petrosal nerve and nerve of the pterygoid canal (Vidian nerve) to the sphenopalatine ganglion From here postganglionic fibers pass via orbital branches of the ganglion to the lacrymal gland, and to the mucous membrane of the soft palate, nasopharynx and pharynx via the palatine nerves Vasodilator fibers also leave the facial by the great superficial petrosal nerve, and entering the cranium are conveyed along the middle meningeal artery and its branches

2) The parasympathetic fibers (secretory and vasodilator) of the *glossopharyngeal nerve* arise from cells of the *inferior salivatory* (*sali-*

[1] This nucleus is actually in the lowest part of the pons

vary) *nucleus* This nucleus lies in the uppermost part of the medulla oblongata immediately below the superior salivatory nucleus, and lateral to the motor nucleus of the glossopharyngeal nerve The autonomic fibers leave the brain with the latter nerve but separate from it again in its tympanic branch (Jacobson's nerve) which joins a twig from the geniculate ganglion (g of facial nerve) and filaments from the internal carotid plexus to form the *tympanic plexus* From this plexus emerges the *small superficial petrosal nerve* through which parasympathetic fibers are continued to the otic ganglion From this ganglion postganglionic fibers are conveyed to the parotid gland via the auriculotemporal nerve The tympanic plexus itself sends filaments to the mucous membrane of the tympanic cavity, the mastoid air-cells, auditory (Eustachian) tube, and the internal ear

3) The *vagus nerve* contains the greater proportion of the fibers of the bulbar outflow They arise from the *dorsal motor nucleus* of the vagus and are distributed through the latter's numerous branches to the thoracic and abdominal viscera (chap 2) Unlike those in the other two cranial nerves, the preganglionic fibers of the vagus connect with ganglion cells situated within the innervated organs Thus, the vagal fibers to the heart connect with ganglion cells in the cardiac walls those to the bronchi with the nerve cells of the intrinsic plexus in the bronchial walls, those to the esophagus, stomach, and intestine form synapses with the ganglion cells of the myenteric plexus of Auerbach and the submucous plexus of Meissner The preganglionic fibers are therefore quite long, the postganglionic very short, and the analogue of the terminal ganglia may be represented as relatively small collections of postganglionic neurons The cells of origin of most of the cardiac fibers of the vagus form a discrete group (cardioinhibitory center) lying alongside the dorsal nucleus of the vagus

SACRAL OUTFLOW

The cells of origin lie in the intermediate columns of the 2nd, 3rd and 4th, and sometimes the 1st sacral segments of the spinal cord The preganglionic fibers emerge in the anterior roots of the corresponding sacral nerves, then leave the spinal nerves again and, proceeding peripherally as the *pelvic nerve* (or *nervus erigens*), on each side, enter into the formation of the pelvic plexus The fibers terminate around ganglion cells lying in close relation to the pelvic organs They carry motor impulses to the walls of the descending colon, rectum, and bladder, inhibitory impulses to the internal anal and visical sphincters and to the uterus, and dilator impulses to the blood vessels of the bladder, rectum, and genitalia

Functions of the Autonomic Nervous System

Because the autonomic nervous system transmits information to the spinal cord and brain from the viscera, and because it provides a significant part of the motor control of all cardiac and smooth muscle, of exocrine glands and possibly of certain endocrine organs, its function may be designated as visceral control It is not, however, the sole means of such control, inasmuch as the digestive tract, for example, has an elaborate pattern of control based upon intrinsic and extrinsic hormonal mechanisms A complete account of the control of digestion, therefore, must include both neural and hormonal components They are alike in that both appear to be "involuntary," operating without conscious awareness of their activity They are nevertheless interdependent, and both are under the influence of the central nervous system

Through its various activities the autonomic system contributes to the maintenance of the constancy of composition of the fluid environment of the body's cells, it serves to combat forces, acting either from within or without, which tend to cause variations in this environment Regulation of the composition of the body fluids and their temperature, quantity, and distribution is effected through the actions of the autonomic nerves upon circulatory, respiratory, excretory, and glandular organs For example, of glandular structures under autonomic influence, the liver, pancreas, and adrenal medulla are of especial importance in the regulation of blood sugar, the sweat glands aid in the control of body temperature The stability of the internal environment (the *milieu interne* of Claude Bernard) which is so characteristic of the healthy body, was designated by Cannon as *homeostasis*. According to Cannon, the essential and particular function of the autonomic system is to contribute to the

internal adjustments upon which this constant state depends

The great majority of the effector organs of the autonomic system are innervated by both sympathetic and parasympathetic divisions, and the effects exerted by the two types of fiber going to a given organ are antagonistic Thus, the heart's action is inhibited by the vagus but augmented by the sympathetic In the intestine the effects of the two systems are reversed, the parasympathetic (vagus nerve to the small bowel and upper half of the colon, and the pelvic nerve to the lower half of the latter) is augmentative, the sympathetic is inhibitory The removal of the effects of one set of fibers, as by section, results, as a rule, in the effects of the other set becoming more prominent, at least temporarily This fact indicates that each type of fiber exerts a constant or tonic action and suggests that the two effects are delicately balanced one against the other Thus, section of the vagus nerves causes an increase in the cardiac rate, and section of the parasympathetic or of the sympathetic fibers to the iris causes, respectively, dilation or constriction of the pupil

Taken as a whole the actions of the sympathetic division including the medulla of the adrenal gland are directed towards strengthening an animal's defenses against the various dangers which beset it, e g , extremes of temperature, deprivation of water, or the attacks of its enemies It has been shown by Cannon et al (1929), however, that the sympathetic system is not indispensable, both ganglıated cords may be completely removed yet the animal remains in good health provided it is kept in the sheltered environment of the laboratory Sympathectomized cats if kept warm and carefully tended will live indefinitely Kittens deprived of their sympathetic trunks grow normally, female cats become pregnant and give birth to young, though the mammary glands do not function normally because of interruption of the sensory nerves supplying the nipples Sympathectomized animals are, however, incapable of arduous work, sugar is not mobilized from the liver on demand, an increase of circulating red blood cells does not occur during excitement or exercise, the usual reactions to cold (elevation of the hairs and vasoconstriction) fail, and epinephrine is not liberated in an emergency They are also less able to withstand oxygen lack or hemorrhage than are normal animals It is evident that the sympathectomized animal could not fend for itself, and, in the struggle for existence, would soon succumb to the hazards of the environment The sympathetic ganglionated chain has also been completely removed in man, thus abolishing all connection of the peripheral sympathetic system with the central nervous system, without producing any serious consequences

Parasympathetic effects, as noted above, are more localized in character It has also been suggested that they are concerned with conservative and restorative processes, in contrast to the sympathetic with processes involving the expenditure of energy Inhibition of the heart, contraction of the pupil for the protection of the eye from intense light and the activities of the digestive tract, through which the energy stores of the body are replenished, are frequently given as examples of acts of conservation presided over by the parasympathetic These apparent differences in the activities of the two divisions have led some (following Gaskell) to speak of the functions of the sympathetic and parasympathetic divisions as *catabolic* and *anabolic*, respectively It is an interesting but essentially a philosophical generalization

Chemical (Humoral) Transmission of Nervous Impulses

In the autonomic nervous system the transmission of impulses from neuron to neuron, and from neuron to effector cell is accomplished by the elaboration of chemical substances by the nerves (sect 1, chap 3) The classical experiments of Otto Loewi, in 1921, first clearly demonstrated the existence of humoral mediators Two frog hearts, arranged in standard fashion for recording contractions, were prepared so that the fluid perfusing one heart could come in contact with the second Upon stimulation of the vagus nerve to the first heart, its rate of contraction was reduced When the fluid perfusing this heart came in contact with the second it also slowed in rate although its vagus had not been stimulated Some substance had obviously been released in the first heart which could be carried by the perfusing fluid to the second Loewi termed this material *Vagusstuffe* By the same technique, stimulation of the cardiac accelerator nerve to the first heart increased its rate and

also the rate of the second heart when the perfusing fluid reached the latter This response Loewi attributed to the release of *Acceleranss-tuffe*

By various techniques, *Vagusstuffe* has been shown to be acetylcholine, and *Acceleranss-tuffe* is norepinephrine In 1933, Sir Henry Dale proposed the terms *cholinergic* and *adrenergic* to describe the nerves which liberated acetylcholine or epinephrine (or related substances) (Fig 9 47) The adjectives are also used to describe agents which cause effects similar to those produced by acetylcholine and epinephrine Dale also proposed that cells stimulated by cholinergic or adrenergic nerves be termed cholinoceptive and adrenoceptive, respectively

CHOLINERGIC TRANSMISSION

Acetylcholine is now recognized as the mediator of impulses at all autonomic ganglia, in the adrenal medulla, at all parasympathetic

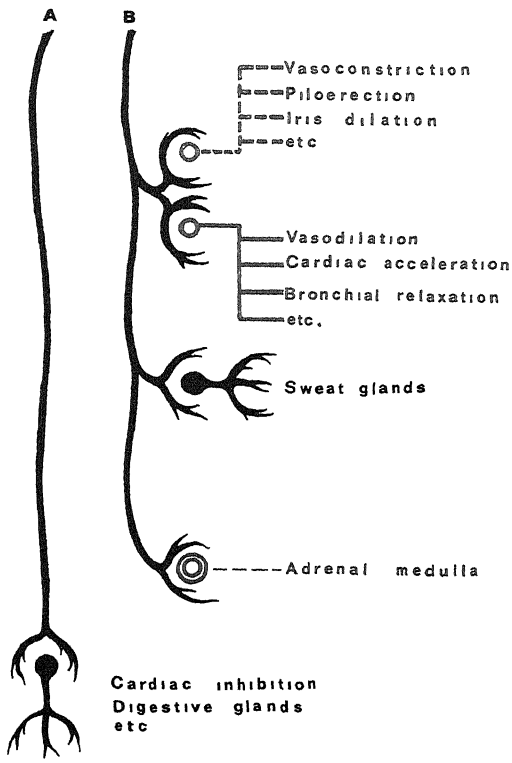


Fig 9 47 Distribution of cholinergic and adrenergic fibers in parasympathetic (*left*) and sympathetic divisions of the autonomic system Cholinergic mechanisms are in black adrenergic neurons are in color

postganglionic terminations, sympathetic postganglionic endings at sweat glands, and motor nerve endings at skeletal muscle In these cholinergic nerve fibers acetylcholine is synthesized (Ritchie and Goldberg, 1970), and retained in an inactive form in morphologically discrete vesicles Upon stimulation of the nerve the acetylcholine is set free and diffuses from the terminations (Paton et al , 1971) The same vesicles contain an enzyme, cholinacetylase, which promotes the synthesis of acetylcholine from choline and acetyl coenzyme A Cholinacetylase is very widely distributed in the nervous tissues, being found in brain, ganglia, and peripheral nerves, it is also present in the placenta, but is absent from liver and kidney For optimal activity this enzyme requires Mg^{++}, K^{+} and Ca^{++} ions (Fig 9 48)

Transmission of excitation from one neuron to another is accomplished by the release of acetylcholine from these synaptic vesicles into the "cleft" between the terminals of the first neuron and the receptor mechanism of the second Calcium ions are required for this release of acetylcholine Fatt and Katz (1952) demonstrated that the presynaptic transmission process is not infinitely graded, but exhibits discrete increments that depend upon the release of acetylcholine molecules in bursts of several thousand molecules each (Fig 9 13, p 9-13) They correlated postsynaptic endplate potentials with the release of these packets or quanta of transmitter molecules

The action of this acetylcholine upon the postsynaptic cell is compatible with the hypothesis that it combines with a receptor substance in the membrane under the presynaptic termina The receptor is believed to be limited to the exterior of the membrane, because Del Castillo and Katz (1955) found that intracellular injection of acetylcholine did not activate the postsynaptic cell Ordinarily the receptor is found only adjacent to the terminal If the presynaptic nerve fiber is cut, however, after some days the postsynaptic cell seems to become hypersensitive to the transmitter substance—in this instance, to acetylcholine When this occurs the area of the membrane sensitive to acetylcholine spreads from the region of the terminals until the entire surface of the muscle cell becomes sensitive With reinnervation, the receptive surface contracts until it is once again limited to the area under the

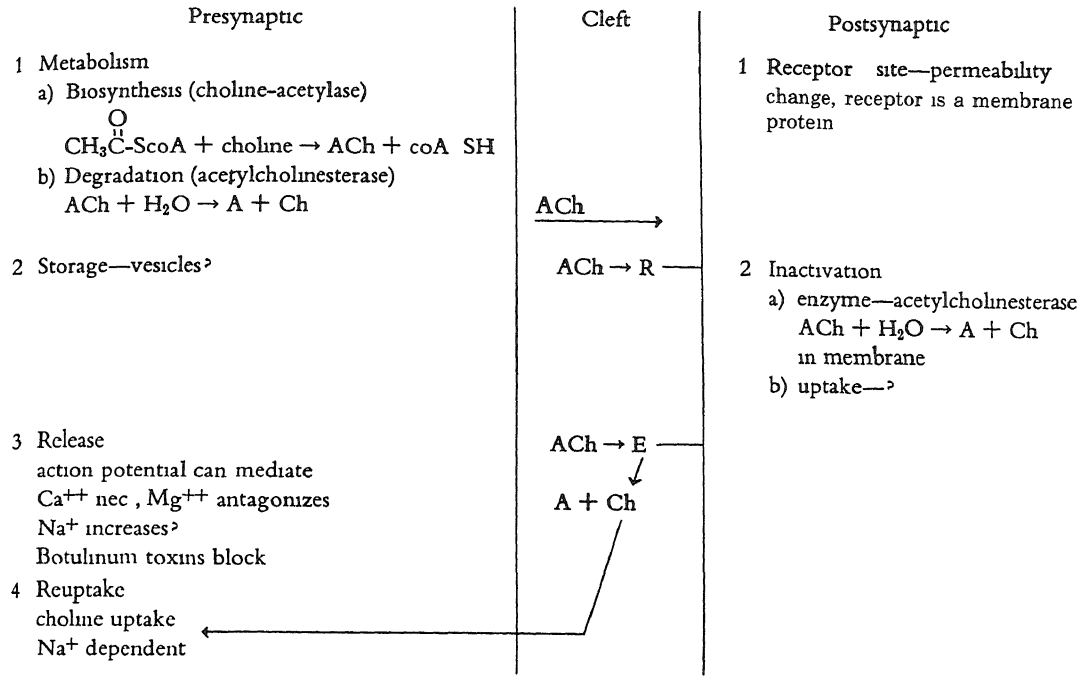


Fig 9 48 Synaptic chemistry of acetylcholine (Kravitz 1967)

terminals (Miledi, 1960) The nature and identity of the receptor substance are unknown

After its release acetylcholine is unstable, it is readily hydrolyzed to choline and acetate ion This hydrolysis occurs almost instantaneously in the presence of *acetylcholinesterase*, an enzyme present at the termination of cholinergic nerves The enzyme is one of a group of hydrolytic enzymes, often referred to as cholinesterases, found in the body Acetylcholinesterase, or *true* or *specific cholinesterase*, has acetylcholine as its preferred or only substrate, and is most effective when the amount of acetylcholine is small The other (pseudo- or nonspecific) cholinesterases will utilize acetylcholine and other choline and noncholine esters as a substrate and are most effective in the presence of large quantities of acetylcholine

Cholinesterases may be inhibited by a variety of substances Much of the knowledge about acetylcholine and nerve impulse transmission has been obtained through the ability of these agents to inhibit acetylcholinesterase Physostigmine (eserine), an alkaloid extracted from the Calabar bean (*Physostigma venenosum*) was the first substance shown to inhibit these enzymes Subsequently many synthetic inhibitors have been prepared, among them neostigmine (Prostigmin), di-isopropylfluorophosphate (DFP) and tetraethylpyrophosphate (TEPP) Physostigmine and neostigmine are reversible inhibitors, the enzyme is inactivated only so long as it is in combination with the inhibitor As the latter is eliminated by the body, the enzyme once more becomes active DFP and TEPP combine irreversibly with the enzyme, permanently inactivating it Hence, for cholinesterase activity to be restored new enzyme must be synthesized

Acetylcholine Liberation during the Discharge of Epinephrine. Stimulation of the sympathetic fibers supplying the adrenal medulla causes an acetylcholine-like substance to appear in the blood of the adrenal vein It has therefore been concluded that acetylcholine acts as a chemical transmitter from the nerve terminals to the medullary cells It will be recalled that the sympathetic fibers ending in the adrenal are preganglionic, the adrenal cell being in fact a postganglionic neuron

Acetylcholine as the Transmitter of

Effects to the Sweat Glands. Although the human sweat glands are innervated by the sympathetic, their behavior to drugs is similar to that of structures supplied by the parasympathetic, they are unaffected by epinephrine, stimulated by pilocarpine, and paralyzed by atropine Dale and Feldberg (1934) showed that excitation of the sympathetic fibers to the foot pads of the cat is followed by sweating and the appearance of acetylcholine in the eserinized fluid perfusing the paw

ADRENERGIC TRANSMISSION

Knowledge of the chemistry of the adrenergic mediator has increased dramatically in recent years Cannon and Bacq in 1931 demonstrated that stimulation of adrenergic nerves liberates into the circulation a substance which they called "sympathin " Although many of the effects of sympathetic stimulation could be reproduced by administration of epinephrine, Cannon and Rosenblueth (1933) were unable to demonstrate a complete qualitative and quantitative identity between the two types of stimulation U S von Euler (1946) then showed that "sympathin" is a compound related to *epinephrine*, namely, dl-(3,4-dihydroxyphenyl) ethanolamine or *norepinephrine*

The confusion arising from much of the original work with the adrenergic mediators was related to the difficulty in separating epinephrine and norepinephrine, and correctly identifying these compounds For example, extracts of the adrenal medulla contain roughly 10% of norepinephrine as a contaminant of what was believed to be pure epinephrine On the other hand, norepinephrine is the principal if not the only chemical mediator at adrenergic postganglionic sympathetic nerve endings (see von Euler, 1971)

Epinephrine and norepinephrine belong to a class of compounds known as catecholamines, synthesized in the body from phenylalanine The series includes a precursor, (3,4-dihydroxyphenyl) alanine (dopa), (3,4-dihydroxyphenyl) ethyl amine ("dopamine"), (3,4-dihydroxyphenyl) ethanolamine (norepinephrine, arterenol); methyl norepinephrine (epinephrine), and another compound known as isopropylarterenol or isoproterenol (see Fig 9 49) These compounds have been studied intensively, so that their pathways of synthesis are largely known, and also the reactions through

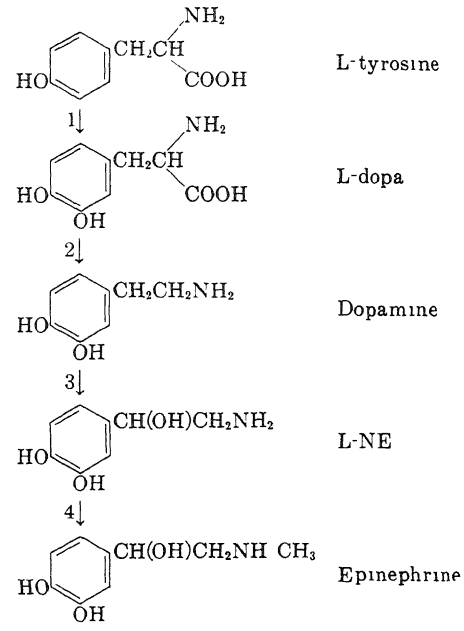


Fig 9 49 Biosynthesis of catecholamines from L-tyrosine (i e hydroxyphenylalanine) The enzyme for the fourth step phenylethanolamine *N*-methyl transferase is present in adrenal medulla and chromaffin tissues but not in typical postganglionic adrenergic endings (Iversen 1967)

which they are inactivated in biological systems (Iversen, 1967)

The synthesis of the adrenergic mediators from phenylalanine can proceed through a variety of pathways There are vesicles present in adrenergic nerve terminations which apparently subserve the function of storage of norepinephrine (Geffen and Livett, 1971), and which contain the enzyme, *dopamine-β-oxidase*, which converts dopamine to norepinephrine. Nerve terminals have an interesting mechanism for conserving norepinephrine, inasmuch as most of the norepinephrine present in these vesicles comes from that which was liberated by the nerve ending, the substance diffusing back into the nerve, and thence into the vesicles (Fig 9 50) Stromblad and Nickerson (1961) demonstrated that both epinephrine and norepinephrine, exogenously administered, could be stored in the tissues, and that the storage was less in chronic sympathetically denervated tissues, suggesting that the site of storage was in the nerve endings

The concept of the mobility of the adre-

nergic mediator out of, and back into, the nerve endings is important for an understanding of various drug effects and the duration of the effects of adrenergic nerve stimulation For, compared to acetylcholine, the adrenergic mediators are inactivated relatively slowly

Inactivation. At least four possible pathways of epinephrine metabolism have been described *1*) An enzyme, *monoamine oxidase*, is present throughout the body which is capable of oxidatively deaminating primary amines Epinephrine and norepinephrine are substrates for this enzyme, although the reaction proceeds rather slowly *2*) Conjugation of epinephrine with glucuronic or sulfuric acid at the hydroxyl groups occurs, primarily in the liver and intestines *3*) The phenolic hydroxyl groups of epinephrine are readily oxidized, forming a quinone which in turn forms a cyclic structure *adrenochrome* *4*) Axelrod (see Axelrod and Kopin, 1969) has demonstrated that the phenolic group in the *meta* position is methylated by the enzyme, O-methyltransferase, in the body (Fig 9 51) This probably is the principal metabolic pathway for circulating epinephrine and norepinephrine It appears, however, that none of these inactivating mechanisms is the one primarily responsible for terminating adrenergic effects Rather, the termination of adrenergic effects results from the diffusion of the mediator from the effector site back into the nerve termination

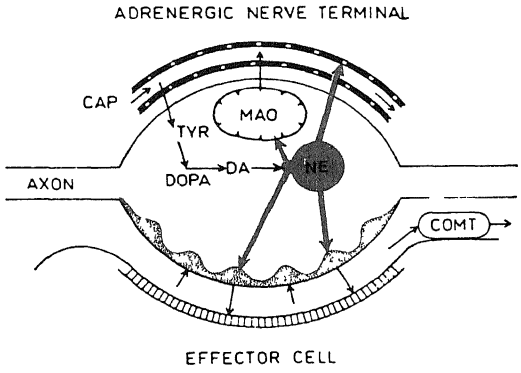


Fig 9 50 Schematic drawing of adrenergic nerve terminal Pathway of synthesis of norepinephrine (*NE*) from tyrosine (*TYR*) is shown together with presence of monoamine oxidase (*MAO*) Norepinephrine is released from storage sites and also re-accumulated into these sites some is oxidized by catechol-*O*-methyl transferase (von Euler 1971)

Receptor Sites The concept of specific receptors, first proposed by Dale in 1906, aids in understanding the rather complex responses to sympathetic stimulation or epinephrine administration Noting that epinephrine caused some smooth muscle to contract, while it relaxed others, he proposed that the former had "motor" type receptors while the latter had "inhibitory" type receptors In 1948 an American pharmacologist, Ahlquist, proposed a modification of this concept that is now universally accepted He utilized five different catecholamines, including epinephrine, norepinephrine, and isoproteranol, in a comparative test of their potency on a variety of physiological reactions He found that these amines had one rank order of potency in evoking vasoconstriction, excitation of uterus and ureters, contraction of nictitating membrane, dilatation of pupil and inhibition of smooth muscle in the intestine But they had a different rank order of potency when tested on functions that included vasodilatation, inhibition of the uterus, and myocardial stimulation From these results he concluded that the effector cells must

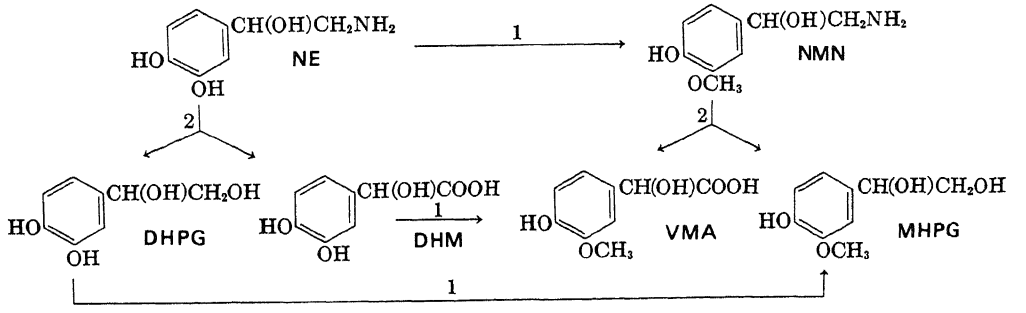


Fig 9 51 Pathways by which catecholamines are inactivated by catechol-*O*-methyl transferase (1) or by monoamine oxidase (2) *NE* is norepinephrine *NMN* is normetanephrine *DHPG* is 3,4-dihydroxyphenylethylglycol DHM is 3 4-dihydroxymandelic acid *MHPG* is 3-methoxy 4-hydroxyphenylethylglycol *VMA* is 3-methoxy, 4-hydroxymandelic acid (Iversen, 1967)

possess two different types of adrenergic receptor mechanisms, one of which is associated with most of the excitatory functions and inhibition of smooth muscle of the intestine The other is associated with most of the inhibitory functions plus an excitatory action upon the heart These hypothetical receptors, given the designation α and β, have been extensively studied by later investigators, with the general conclusions shown in Table 9 2 Furchgott (1954) has presented evidence which indicates that with vascular smooth muscle, at least, individual cells may contain both α and β *receptors*

Norepinephrine appears to stimulate only α-receptors, isoproterenol only β-receptors, and epinephrine is capable of stimulating both Epinephrine is 2 to 10 times more potent than norepinephrine in stimulating the α receptors Isoproterenol is 2 to 10 times more potent than epinephrine in stimulating the β receptors and 100 or more times more active than norepinephrine Therefore, although epinephrine is the more potent agent as a vasoconstrictor, norepinephrine may produce a greater rise in blood pressure for a given dose, since it has little or no effect on vasodilator fibers

Table 9 2
Functions of α and β adrenergic receptors

α Receptors
Vasoconstriction, especially in vessels of skin, mucosa, gut, salivary glands, lungs, cerebral circulation
Stimulation of contraction of radial muscle of iris (pupillodilation), of nictitating membrane, of pilomotor muscles, of muscle of uterus, and of ureter
Stimulation of glycogenolysis
Inhibition of insulin secretion
Inhibition of smooth muscle of intestine
β Receptors
Vasodilatation in coronary circulation and in skeletal muscle, and possibly in the liver
Relaxation of muscle of bronchi, and of uterus
Stimulation of lipolysis
Stimulation of insulin secretion
Stimulation of heart, with cardioacceleration, increase in contractility of muscle of atria and ventricles

Higher Levels of Visceral Control Systems

Both afferent and efferent components of the visceral control systems can be demonstrated in levels above the segmental or reflex mechanisms of spinal cord and cranial nerves The cell bodies of the interneuronal (preganglionic) neurons lie in a column of cells that is continuous throughout thoracic and lumbar segments of the cord, can be identified but is less conspicuous in sacral segments, and is grossly interrupted in the brainstem where it has the form of four isolated cell groups (see preceding chapter) In the spinal cord it is called the intermediolateral cell column of the central grey substance, because of its location and relationships, in the brainstem each of the four cell groups is designated as a "nucleus"—the Edinger-Westphal nucleus of nerve III, the salivatory nuclei of nerves VII, and IX, and the dorsal motor nucleus of the vagus (nerve X)

If there is any processing of information, or modulation of control, within autonomic ganglia it is of only a simple nature, and can involve only the terminations of preganglionic upon postganglionic neurons, inasmuch as afferent fibers do not send terminals to cells in the ganglia Synaptic functions within ganglia seem to be stereotyped A functional significance is attributed, however, to the difference in innervation ratios in the two divisions of the autonomic motor system In the craniosacral portion each preganglionic neuron innervates only a limited number of postganglionic cells, and control tends to be discrete rather than diffuse In the thoracolumbar portion, by contrast, fibers from white rami may terminate in a given ganglion, or also run through it or even through several ganglia, with terminations upon as many as 30 postganglionic cells This means that responses tend to be diffuse or generalized This tendency is reinforced by the circumstance that one of the principal collections of thoracolumbar postganglionic cells is the medulla of the adrenal gland By discharging their "transmitter," which is epinephrine with smaller percentage of norepinephrine, directly into the circulation they assure the generalized character of any significant sympathetic activation

A lack of data regarding what happens in the intermediolateral cell column makes it impossible to discuss the nature of control at this level On the other hand, there are plentiful data regarding stimulation of autonomic

otor systems both directly and reflexly thin brainstem, diencephalon, and even cerellum and cerebral cortex Neurons that can activated at these levels must be regarded connector, internuncial neurons, but they e also computational in that they are responle for naturally occurring control responses ie earliest experiments upon these higher echanisms were intended to disclose the lotion of a mechanism that produces glycosa after surgical injury to the medulla oblonta, the *sucre piqûre* of Claude Bernard They l to the concept that the medulla contains a eflex" center for control of epinephrine setion In later experiments electrical stimuion of the floor of the 4th ventricle was ind to produce a rise in blood pressure, aceration of the heart, peripheral vasoconiction, and secretion of epinephrine, and it came clear that autonomic controls are jor functions of the brainstem These sysns are now classified as the so-called medully cardiac center, vasopressor center, or vasepressor center (please see sect 8, chaps ind 9) Similar studies related to control of lmonary ventilation involve "automatic" ; not autonomic mechanisms, as discussed section 6, chapter 4

Just before the first World War, two Viense investigators, Karplus and Kreidl (1910), covered that they could evoke a surprising riety of sympathetic responses by stimuing the base of the brain in animals where y had exposed the brain by removal of one it Ranson (1936, 1937) and his associates er repeated their experiments with the preion of the sterotaxic method, and described inges in heart rate, blood pressure, respiran, secretion of salivary and other glands, as ll as changes in the motility of stomach, vel, and bladder after stimulation within hypothalamus The most active sites were the lateral hypothalamus, and some of the ponses obtained were in the nature of inhiion rather than stimulation (Ingram, 1939) m their results and those of other laboratos came almost innumerable attempts to asn the several portions of the hypothalamus either the craniosacral or thoracolumbar systems Many authors have concluded that the rostral hypothalamus tends to be craniosacral in its orientation, with the more caudal part thoracolumbar As a generalization, however, it is more nearly correct to say that the lateral hypothalamus tends to be excitatory and the medial hypothalamus inhibitory, no matter what kind of control system is under consideration Moreover, from an electrode at any given locus in the hypothalamus one may obtain either a rise or a fall in blood pressure and heart rate, depending upon the frequency and strength of the electrical stimulation These observations, together with the disclosure that the hypothalamus participates in predictable ways in control systems having obvious functions suggests that the goal of identifying discrete "centers" for sympathetic as opposed to parasympathetic actions is not in keeping with the organization of this portion of the brain

Although the hypothalamus is sometimes said to be the level of highest control of autonomic functions, the "head ganglion" of the autonomic system, it is not the highest level from which autonomic responses can be evoked There are reports of alterations in blood pressure and heart rate, as well as in peripheral distribution of blood flow, following stimulation of certain regions of the cerebral cortex, and even after cerebellar stimulation Neither the nature of the responses nor the location of the active sites lends itself to any generalization as to the significance of these results One possibility is that, inasmuch as Pavlovian conditioning in general requires cortical function, and inasmuch as autonomic responses readily show conditioning, this may be the significance of these cerebral locations One cannot say, however, whether any particular lesion or ablation of the cerebral cortex will reduce or abolish any one of these conditioned responses Another interesting possibility has followed the discovery that many of the regions from which autonomic responses can be evoked belong to what is known as the "limbic system" (MacLean, 1970), the portion of the brain believed to be responsible for emotional experience (chap 8, p 9-156)

4 / CONTROL OF SKELETAL MUSCLE

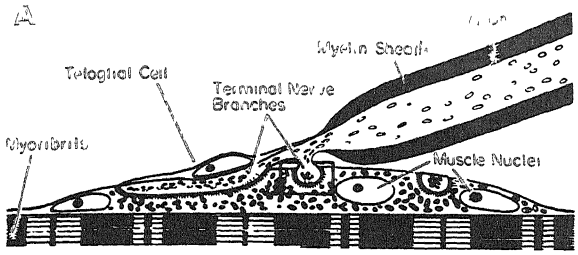
The process that ultimately evokes shortening of skeletal muscle via the actomyosin complex is believed to be a release of calcium ions from the membrane of the muscle fiber In mammalian striated muscles this calcium is stored in what is known as the *triad*, a structure adjacent to the transverse (or T) tubules of the sarcoplasmic reticulum, and also adjacent to the cisternae that run parallel to the fibrils The lumen of the transverse tubule is continuous with the extracellular compartment surrounding the muscle cell, and the membrane lining the tubule is likewise a projection of the sarcolemmal membrane When microelectrodes are placed on the sarcolemmal membrane in the region of origin of a transverse tubule, the application of an electrical stimulus produces a contraction of only the two adjacent half-sarcomeres on each side of the tubule (Huxley and Straub, 1958, Huxley and Taylor, 1958) Apparently the excitatory process initiated in the sarcolemma is propagated into the muscle via the transverse tubule, where it brings about a release of calcium ions from the triad, and hence the initiation of the contraction process (see Peachey, 1965, Winegrad, 1965)

If the ultimate process is taken to lie in the triad and sarcoplasmic reticulum, at least two other stages of excitation and transmission are interposed between this final level and the nerve impulse of a motor nerve Just above the level of the reticulum is the membrane of the muscle cell It is this membrane that may or may not be activated by natural or artificial stimulation It is also this membrane that gives the contraction process some of its classic characteristics, including the all-or-nothing phenomenon The membrane may undergo local changes that electrically have the characteristics of electrotonus, the local excitatory state (LES) described in chapter 1 When this local state becomes intense enough in a short enough time course, a disturbance analogous to a nerve impulse is spread over the surface of the fiber, and throughout its sarcoplasmic reticulum

At the next higher level of control is the nerve ending on the muscle, which typically has the form of a motor end-plate (Fig 9 52) The nerve impulse reaches the ending, and there sets up processes similar to those responsible for synaptic transmission in autonomic ganglia The onset and course of these processes may be followed by recording of potential changes, but the processes per se are physicochemical, and involve the release of acetylcholine from vesicles in the nerve ending, the diffusion of the acetylcholine across the neuromuscular "cleft," and some unknown interaction of the acetylcholine with a receptor substance present on the sarcolemmal membrane only in the region of the motor end-plate The acetylcholine so released is inactivated almost at once by acetylcholinesterase, but if the quantity of acetylcholine released and the time course of its release were adequate, the combination of the transmitter and the receptor substance alters the membrane potential of the sarcolemma sufficient to initiate the wave of excitation noted above The electrical sign of this excitation is a reversal in polarity of the muscle's membrane potential, As in nerve, this *action potential* is not identical with the wave of excitation that induces it, but the more fundamental reaction known as a *nerve impulse* in nerve has no comparable brief name when it occurs in muscle

Central Neural Control of Muscle

All striated or skeletal muscles of higher animals are controlled by neurons located in the motor columns of spinal cord and brainstem, that send their axons to the periphery in spinal or cranial nerves In the spinal cord these neurons are collected in the anterior grey columns into groups innervating the individual muscles or muscle groups (chap 1) In the brainstem they are present in two distinct columns One innervates muscles derived embryologically from myotomes, and includes the motor nuclei of the oculomotor, trochlear, abducens, and hypoglossal nerves The other innervates striated muscles derived from the



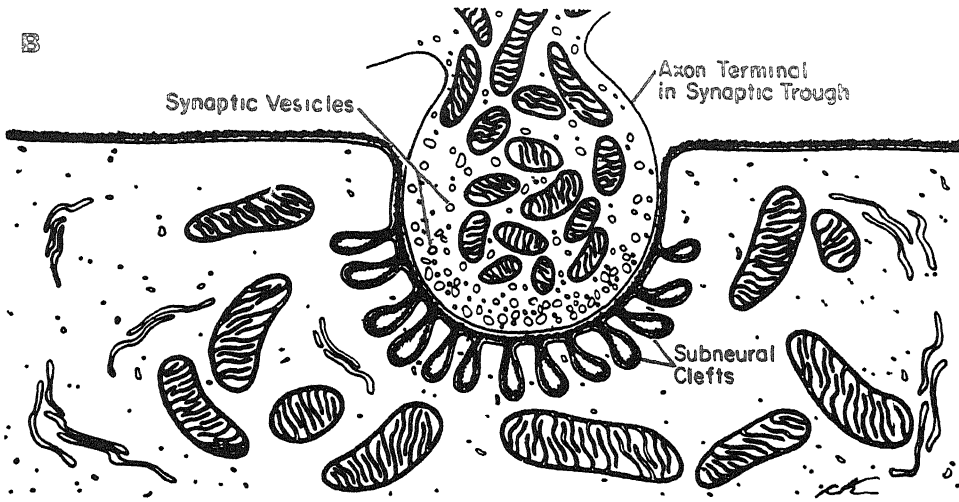


Fig 9 52 Motor end plate for neuromuscular transmission in striated muscle *A* Diagram of microscopic appearance *B* diagram based on electronmicroscopic data (Bloom and Fawcett 1968)

embryonic branchial arches, and includes motor nuclei of the trigeminal and facial nerves, in addition to the nucleus ambiguus that gives rise to motor fibers in the glossopharyngeal, vagus, and accessory nerves to supply striated muscles of the pharynx and larynx

It is not possible to describe the control of skeletal muscle by considering only the motor neurons They are continuously under the influence of other neurons, including mainly the afferent (receptor) neurons and the connector (internuncial) neurons of that level of the neuroaxis They are also affected by nerve impulses that come from the opposite side of the spinal cord or brain, and by impulses from lower and especially from higher mechanisms In spite of this complexity, many of the properties of the motor control systems can be disclosed by study of the segmental mechanisms responsible for what is known as *reflex action*

CHARACTERISTICS OF REFLEX ACTIVITY

The anatomical basis of reflex action is the reflex arc (Fig 9 53), which has three basic components *a*) An afferent limb is composed of the receptor organ, which upon excitation, gives rise to nerve impulses and the neuron whose processes transmit the impulse to the central nervous system, it may consist of the spinal reflex arc, the cell bodies of the afferent neurons are situated in the posterior root ganglia *b*) An efferent limb is constituted of a motor neuron which conducts impulses from the central nervous system to an effector organ-muscle or gland In the case of motor spinal reflex arcs, the axons of the efferent neurons leave the cord by the anterior nerve roots and travel in the peripheral nerves, their cell bodies are situated in the anterior horns *c*) A "center" is situated in the gray matter of the central nervous system, it may consist of the cell body and dendrites of the efferent neuron

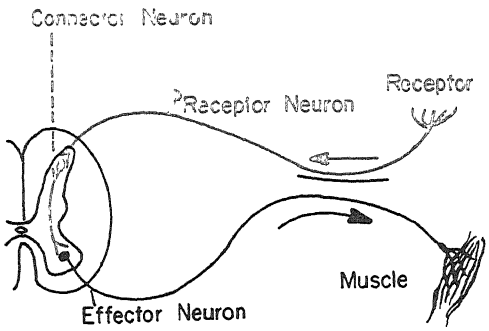


Fig 9 53 Diagram of a simple reflex arc

and its junction (synapse) with the central process of the afferent neuron

The afferent and efferent limbs may connect directly in the center, but usually one or more nerve cells may be interposed between the two These are spoken of as connector, internuncial, or intercalated neurons (Fig 9 53) The stretch reflex as shown by Lloyd (1943, 1944) can be carried out through a reflex arc of only two neurons, but the great majority of the spinal reflex arcs in higher animals consist of several neurons, and in most reflexes each afferent neuron makes connection through collateral branches and internuncial neurons with a large number of motoneurons

Injury, leading to loss of function of any one part of the reflex arc, is sufficient to alter the function of the whole

The scientific study of reflexes may be said to have begun early in the 1800's when Bell (1811) in England and Magendie (1822a and b) in France established the functions of the dorsal and ventral roots of the spinal cord (see Fulton and Wilson, 1966) By the beginning of the present century, Sherrington (1906) at London, Liverpool, and then at Oxford had begun a detailed, quantitative investigation of the properties of extensor and flexor reflexes At about the same time the Spanish anatomist, Ramón y Cajal (1909, 1911) presented histological evidence in support of the neuron "doctrine," asserting that sensory and motor neurons are contiguous but not continuous But neurophysiological investigations of reflex activity did not penetrate to the synaptic level until the introduction of electrophysiological methods The use of such methods was pioneered in the years 1920 to 1926 by Erlanger and Gasser (see 1937), who made extensive employment of the cathode ray oscilloscope for study of electrical responses of nerve and spinal cord

At this time electrophysiological methods were also introduced by Adrian (see 1935) for the investigation of receptor responses With Bronk (Adrian and Bronk, 1929), he carried out one of the most important of the early studies of how information is coded in sensory nerves Some years later the introduction of glass microelectrodes by Ling and Gerard (1949) enabled workers such as Eccles (1953) to examine the events of excitation and inhibition at the neuronal level in the spinal cord

Classification of Reflexes. Reflexes may be classified as segmental, intersegmental, and suprasegmental on the basis of their connections in the spinal cord Those reflexes whose arcs pass through only one segment of the spinal cord include, for example, the patellar reflex or knee jerk Intersegmental reflexes involve ascending or descending pathways, thus, pinching the hind limb of a cat may lead to flexion of a contralateral forelimb Suprasegmental reflexes involve interactions between nuclei above the cord and the segments of the cord itself Such reflexes are exemplified by movements of the head causing extension of the limbs in a spinal animal The principal reflexes utilized for investigating the properties of the synapse have been the segmental

Another classification of reflexes has proved useful in clinical work Reflexes are classified into four categories superficial, deep, visceral, and pathological Superficial reflexes are those elicited from mucous membranes or skin, i e, the surface of the body Examples of such reflexes are the corneal reflex, or blinking of the eye when the cornea is touched, and the plantar reflex, a flexion of the toes when the sole of the foot is stroked Deep reflexes are stretch reflexes in which extension of a muscle is obtained by striking the tendon of a given muscle The patellar reflex (knee jerk), triceps and biceps reflexes are examples of this type of reflex Visceral reflexes are elicited from deep-lying structures examples of such reflexes are the pupillary reflex, i e, constriction of the pupil when light is shined in the eye, and the carotid sinus reflex, in which pressure on the carotid sinus causes a fall in blood pressure Finally, the pathological reflexes consist of those which are present only as an abnormality An example of this type is the Babinski reflex which consists of a spreading of

the toes when the plantar surface of the foot is stroked This reflex may be evoked from humans when injuries to the upper motor pathways of the spinal cord are present

The Reflex Preparation. Two animal preparations have usually been employed to study the properties of spinal reflexes, the spinal and the decerebrate animal The *spinal animal* is prepared by sectioning the cord at any level above the level to be studied If the section is higher than the lower cervical region, artificial respiration is necessary because the motor neurons of the phrenic nerve are cut off from necessary medullary control (sect 6, chap 4) Immediately after the transection, the spinal animal is in a condition called *spinal shock* in which there is complete absence of reflexes below the level of the section Within an hour or two there is a return of the flexion and extension reflexes in the cat and dog, in the human the return may take weeks or more The blood pressure of the spinal animal tends to fall immediately after the section, but if the animal is properly maintained, oxygenation is satisfactory, acid-base balance of the blood is achieved, and the blood pressure returns to normal values

A *decerebrate* animal is prepared by transecting the brainstem through the mesencephalon, usually at a level between superior and inferior colliculi As soon as it recovers from the anesthetic used during the operation, such an animal shows striking exaggeration of stretch reflexes, particularly in extensor and other "antigravity" muscles The limbs are rigid, with hyperextension of the back (opisthotonus) and tail, and tight closure of the jaws The usefulness of either a spinal or a decerebrate preparation is that it permits the study of reflexes without anesthesia that would depress reflex activity Furthermore, by comparing the two types of preparation one can learn something about functional levels in the spinal cord and brainstem, especially when sections are made at varied levels in the spinal cord, and at different levels through the brainstem

Sherrington (1906) studied mainly the spinal reflexes of muscles acting upon the lower limbs of cats (Table 9 3) The lower the plane of the section through the spinal cord the more easily the preparation can be maintained for study His usual procedure involved the following steps general anesthesia using ether, decerebration followed by withdrawal of the anesthesia, exposure of the lower thoracic spinal cord, followed by transection of the cord, exposure of sacral cord, with transection In this

Table 9 3

Typical reflex responses of a dog or cat with a section of the central neural axis classified by level and stage of recovery for flexor and extensor muscles

Operation	Flexor Muscles	Extensor Muscles
Low spinal section i e , transection at or about level of last rib		
i) "Acute spinal," i e , some minutes or hours after transection	Flexor reflex easily evoked Tendon-jerk present ('pluck reflex') Crossed inhibition	Tendon-jerk present
ii) "Chronic spinal" i e , some days, weeks, or months after transection	Flexor reflex easily evoked Tendon-jerk present Crossed inhibition	Tendon-jerk present Imperfect stretch reflex (postural) Extensor thrust Crossed extensor reflex
Decerebrate i e , transection between anterior and posterior colliculi some minutes or hours previously	Flexor reflex less readily evoked, but once evoked has considerable after-discharge Tendon-jerk Crossed inhibition	Tendon-jerk with marked tonic after-discharge Stretch reflex exaggerated and chief factor in decerebrate rigidity Crossed extensor reflex easily obtained with long tonic after-discharge

Creed et al , 1932

fashion the lumbar elargement of the cord, the site of reflex connections for the lower limbs, was left in situ but deprived of influences from both higher and lower levels The decerebration permitted study of the animal without pain even in the absence of anesthesia

The activity of spinal motor neurons was displayed by using a muscle as an indicator The tendon of the muscle was attached to a torsion wire myograph, with a pointer writing upon the smoked drum of a kymograph Other muscles that might pull upon the myograph were either denervated or sectioned

Types of Reflex Studied. Many experiments involved the *flexion reflex*, by definition a reflex contraction of a flexor muscle such as tibialis anticus This is the reflex that withdraws the foot when an animal steps upon a sharp or hot surface It includes contraction of the appropriate flexor muscles, inhibition of the opposing extensor muscles of that limb, activation of extensor muscles of the opposite limb, and changes in tone of muscles in the other two limbs as adjustments to the altered center of gravity of the body

In other experiments the *stretch reflex* was analyzed Its latency is so short that Sherrington at first was not certain it was a true reflex It is a quick, brief, and sometimes powerful contraction to a sharp stretch of a muscle and its tendon, and includes the knee jerk, ankle jerk, biceps or triceps jerk as examples It may be, however, a more prolonged contraction following a continued stretching of the muscle, especially in the case of an extensor muscle in a decerebrate preparation

Although the terms are sometimes confused, it is important to understand that the flexor reflex and stretch reflex are not opposites of one another For example, both flexor and extensor muscles may exhibit stretch reflexes, as do the biceps and triceps muscles in the arm The fact that flexion reflexes are necessarily studied in flexor muscles, while stretch reflexes have been studied mainly in extensor muscles, is responsible for the confusion Flexion and stretch reflexes may be compared or contrasted, but one should not regard the latter as necessarily an extensor reflex

Stimulation used to evoke a flexor reflex is usually an alternating electrical current applied to a cutaneous, sensory nerve More natural stimuli such as pinching of a toe pad, or any other usually painful procedure may be utilized For stretch reflexes the muscle is pulled, briefly for a *phasic* response or continuously for a *static* response Sherrington fastened one end of the muscle to the myograph and then let the table carrying the animal fall away from it (see Creed et al , 1932) A more modern device has the animal in a fixed position on a table that can be elevated by means of a lever and cam, so that the muscle pulls upon a strain guage

FLEXION REFLEX

The significance of Sherrington's study of flexion reflexes (see Sherrington, 1906, Creed et al , 1932) lies not so much in what it revealed as to the character of the reflex response, but rather in the insight he provided into the nature of mechanisms lying within the spinal cord The reflex, itself, has the following characteristics

1) From a muscle such as tibialis anticus in cat, a flexor of the ankle, a reflex contraction can be elicited following stimulation of any of the cutaneous nerves of that leg

2) A stimulus in the form of a single shock gives a single twitch of the muscle

3) Repeated stimulation may give a reflex tetanus.

4) The magnitude of the responses obtained from the several cutaneous nerves differs (Table 9 4)

5) The strength of the response varies with the intensity of the stimulation (Fig 9 54)

6) If two afferent nerves are stimulated simultaneously with weak shocks, the strength of the response may be larger than expected, exceeding the apparent arithmetical sum of the two individual responses This is *spatial summation*

7) If two afferent nerves are stimulated with strong shocks, the resulting response may be less than expected, less than the apparent arithmetical sum of the individual responses This is known as *occlusion*

8) Time required for transmission of the excitation through the spinal cord indicates that more than one synapse is traversed, that is, the flexion reflex involves interneuronal activity

In considering the implications of these results it is necessary to relate all of them to the last one, to the presence of connector or internuncial neurons and more than one synapse

Table 9 4

Strength of contraction of *M Tibialis Anticus* (flexor of the foot) in response to stimulation of various cutaneous, sensory nerves Upon direct stimulation of the motor nerve the maximal force of contraction of the muscle was 2,160 gm

Afferent Nerve Stimulated	Tension of Maximal Reflex Tetanus in Grams	Reflex Tension Expressed as Percentage of Maximal Motor Tetanus
Internal saphenous	800	32
Superficial obturator	165	6 7
Deep obturator	400	16
Nerve to quadriceps and sartorius	1190	44
Musculocutaneous branch of peroneal	1700	69
External plantar	1240	50
Internal plantar	1330	54
Small sciatic	680	28
Hamstring	565	23
Nerve of sural triceps	300 (rather low)	12
Total	8370	

Creed et al , 1932

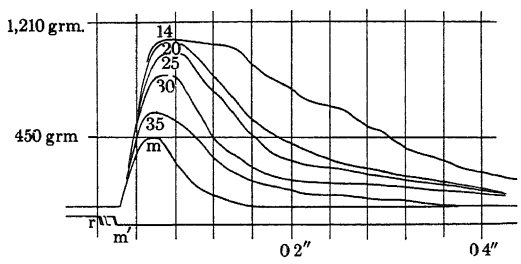


Fig 9 54 Myograms of flexor reflex of tibialis anticus muscle of spinal cat m was recorded after a single shock was applied to the motor nerve the other curves were obtained by increasing the intensity of a single shock applied to the ipsilateral popliteal (a sensory) nerve (Sherrington 1906)

Sherrington inferred that activity brought into the spinal cord via any one afferent nerve is distributed eventually to a variety of collections or "pools" of motor neurons (Fig 9 55) Furthermore, any given group or pool may receive activity from a variety of afferent nerves Finally, the influence of the afferent fibers (or better, the connector fibers) upon motor neurons is not all-or-nothing in magnitude, but is graded such that it is sometimes too weak to be effective, and sometimes reaches a maximum that cannot be surpassed no matter how much stronger the stimulus is made This graded, nonpropagated influence is the central excitatory state (CES) referred to in chapter 1

The hypothesis of the CES explains the first seven observations listed above Tetanus, for example, follows accumulation of CES as a result of the arrival of successive impulses upon motor neurons Any given motor neuron may therefore discharge more impulses and at a higher frequency, and neurons at first not active will be stimulated by accumulation of CES at their membranes Observation 6, spatial summation, is interpreted to mean that each of the weak shocks had a subthreshold effect upon neurons common to the two pools, consequently, when the two weak shocks were applied simultaneously to the two afferent nerves they stimulated neurons that had not discharged impulses following either of the weak shocks alone These particular neurons are said to belong to the *subliminal fringe* of neurons in each neuron pool Similarly, in the experiment showing occlusion, the maximal stimulation of the motor neuron pools supplied by the two afferent nerves included a significant number of neurons that were made active by either stimulus when it was applied alone The result of simultaneous stimulation, therefore, could not be as large as the apparent sum of the two individual stimulations Neurons so affected by two afferent stimulations are said to lie in the region of *central overlap* of the two pools, and the afferent impulses are said to show *convergence* upon these neurons Convergence is also required for explanation of spatial summation, as above

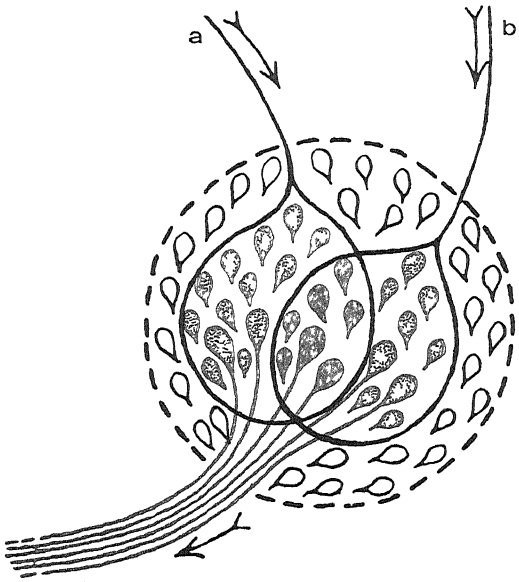


Fig 9 55 Diagram that illustrates hypothetical relationship between excitatory afferent fibers (*a b*) as they end on motor neurons in two pools (enclosed by schematic afferent axons) Motor neurons in the portion common to the two terminal collections can be excited via either afferent This pattern explains the results listed in Table 9 3 (Creed et al 1932)

STRETCH OR MYOTATIC REFLEX

The adequate stimulus which elicits this reflex is a stretch of the muscle If the stretch is performed rapidly a nearly synchronous volley of impulses enters the cord, the reflex response is abrupt and occurs with only the briefest central delay The prototype of these reflexes is the patellar reflex or knee jerk in which a sharp tap on the patellar tendon produces a contraction of the quadriceps muscle The receptors involved in this reflex are the annulospiral endings of the muscle spindle The reflex shows no *after-discharge*, that is the response ceases immediately after the stimulus terminates The measurement of central latency is difficult to achieve by myographic methods since the latency is so short and the input volley not quite synchronous The electrophysiological method to be described below gives a period of about 0 5 to 0 9 msec for the central latency of the patellar reflex

In addition to investigating the responses to a rapid stretch of muscle, it is possible to study the reflex response to a continuous stretch of extensor muscles (Fig 9 56) Since two types of response exist to stretching extensor muscles, the response to a quick stretch is termed the phasic response, whereas the response to continuing stretch is called the tonic response Two components of the tonic reflex response can be observed in the myographic record (Fig 9 56), there is an initial, rapidly rising component which is phasic, and then a sustained, slowly-falling tension which is the tonic component The true course of the reflex, however, can only be determined by excluding the passive tension developed by the muscle This is done by cutting the nerve to the muscle and repeating the experiment A curve of passively developed tension (Fig 9 56) is obtained which can be subtracted from the tension developed when the nerve was intact

The tonic component of the stretch reflex is clearly important in maintaining posture Extension of the muscle, as by swaying away from the center of gravity, activates the stretch reflex and restores the muscle to its initial length The concept that standing may be reflex in nature arose from observations on decerebrate animals in which the limbs are hyperextended Here "reflex standing" appears to be taking place The extensor muscles are reflexly activated by any slight stretch of the muscle Other factors play a role such as a release from inhibiting influences from higher centers but the reflex nature of the rigidity is demonstrable by sectioning the dorsal roots of a single limb, the rigidity then disappears

Another fundamental phenomenon can be observed in these experiments Stretching of antagonist muscles such as the semitendinosus

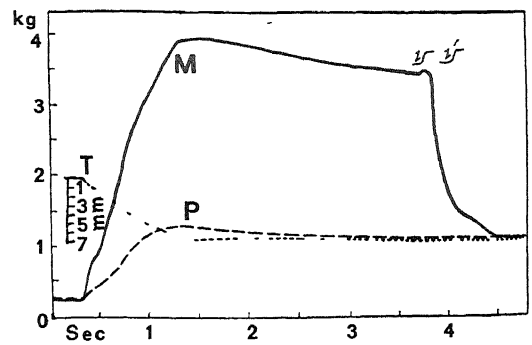


Fig 9 56 Experiment that proved that stretch reflex is truly a neural reflex When muscle was stretched by fall of table (*T*) the contraction of the muscle with normal innervation (*M*) exceeded the elastic resistance of the denervated muscle (*P*) by the difference between *tracings M* and *P* Between *i* and *i'* a stimulation of an ipsilateral nerve inhibited the reflex contractions (Creed et al 1932)

or biceps femoris causes a decrease in reflex tension in quadriceps (Fig 9 57) Indeed, stretching of the biceps causes complete disappearance of the myotatic reflex evoked by sustained stretch to the level given by passive tension alone (curve P) (Fig 9 57) This depression of reflex activity caused by stimulation of specific nerves, in this case the ipsilateral antagonists, is known as *inhibition* and will be studied in greater detail later Inhibition does not simply consist of a lack of reflex activity or a capturing of the motor pathways by the inhibiting nerve It will be shown later that the inhibiting neurons cause a change in the membrane potential of the motoneuron tending to stabilize the motoneuron against responding reflexly

As one might expect, the stretch reflex can be interrupted by section of the dorsal or ventral roots, by division of the nerve to the muscle, or by destruction of the related portion of the spinal cord The afferent fibers in the dorsal roots are large, myelinated, and rapidly conducting fibers of the A group The muscle fibers upon which the motor fibers end are predominantly of the red variety, these particular fibers are responsible for maintenance of posture, and are more plentifully supplied with large diameter motor fibers than are the pale fibers of flexor muscles The name, myotatic unit, is given to the stretch receptor, afferent fiber, reflex connection in the spinal cord, motor neuron, and striated muscle fiber

The principle of *reciprocal innervation* is one of the broad generalizations about spinal reflex activity that resulted from Sherrington's early work on excitation and inhibition The principle states that the flexors and extensor reflexes of the same limb cannot be simultaneously active, if a flexor reflex is evoked, the extensor is inhibited and vice versa For instance, the afferent nerves evoking an extensor reflex have connections going to the flexor motoneurons of the same limb causing them to be inhibited (Fig 9 58) When a flexor reflex is elicited, its connections on the extensor motoneurons cause the extensor reflexes to be inhibited The extent of inhibition in all such cases is proportional to the reflex activity taking place so that a primitive form of organization is taking place at the segmental level and on the limb of one side But at the same time the evocation of a reflex on one side causes opposite effects on the motoneuron pools of the opposite limb Thus, when a flexor reflex is obtained on one side of the animal, the flexors of the opposite limb are inhibited and the extensors are facilitated Two examples of such integrated contralateral activity are the crossed extensor reflex and Phillipson's reflex The *crossed extensor reflex* accompanies a flexion reflex and is obtained by pinching the limb of a spinal or decerebrate animal (Fig 9 59) The pinched limb withdraws but the contralateral limb extends Similarly *Phillipson's reflex* is obtained from the extended limb of a spinal or decerebrate animal, flexing the limb forcibly produces the *clasp-knife reflex* and causes the opposite limb to extend (see Sherrington, 1906) Reciprocal innervation is very useful in more elaborate activities such as walking where it is obviously necessary to have complete synchronization of flexor and extensor activity The presence of reciprocal innervation ensures that accidental evocation of simultaneous reflex antagonist

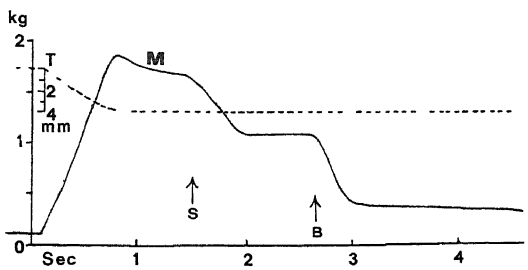


Fig 9 57 Reflex response of quadriceps to stretch (signalled by fall of table *T*) At *arrow S* a stretch was applied to semitendinosus followed at *arrow B* by stretch applied also to biceps The graph shows that stretch applied to the opposing muscles the flexors of the knee inhibited the reflex response in the extensor quadriceps (Creed et al 1932)

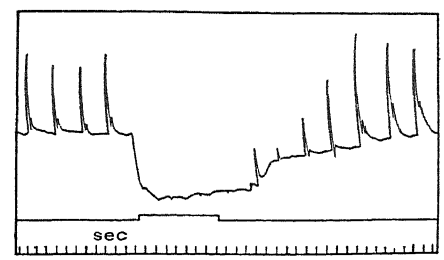


Fig 9 58 Inhibition of knee jerk during central stimulation of a portion of the sciatic nerve Upper tracing is record from myograph middle tracing signals period of central stimulation bottom tracing is time in seconds (Sherrington 1906)

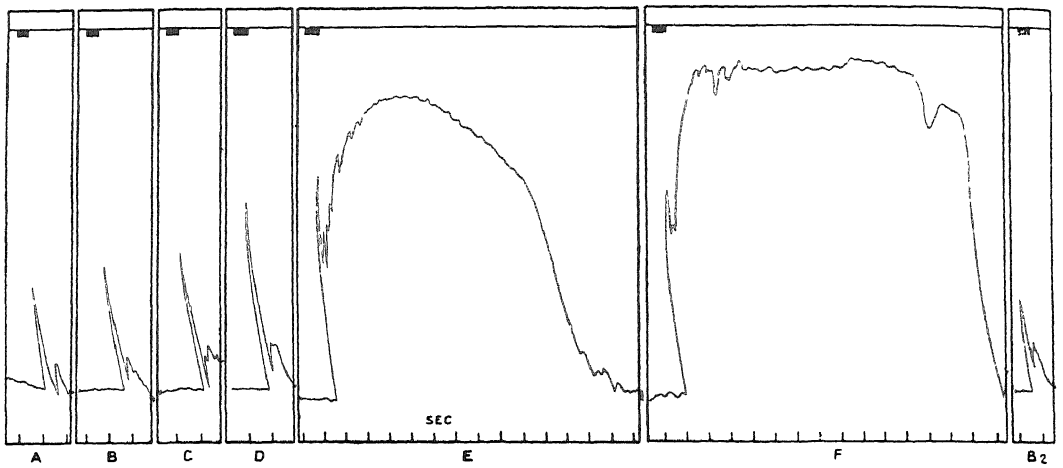


Fig 9 59 Crossed extensor reflex strength and duration of contraction as strength of stimulus applied to contralateral cutaneous nerve was increased (*tracings A–F* B_2 is a repeat of stimulus in *tracing B*) In each instance the stimulus was applied in the form of 12 shocks in ½ sec (Sherrington 1906)

activity does not occur, when one reflex is active, the other is inhibited

Electrophysiology of Spinal Reflexes

Sherrington knew that the latent period of the flexion reflex depends upon the strength of the stimulation, and has a minimal value of about 11 msec (Fig 9 60) To measure accurately the much shorter latencies of the stretch reflex required the development of more sensitive apparatus, including electronic stimulation and recording Electrical potentials recorded when nerve impulses are carried by the ventral roots are referred to as reflex responses, as were myographic recordings of tension If a nerve containing sensory fibers from a muscle is stimulated centrally so that impulses are sent into the spinal cord, a recording of activity in the ventral root will show two waves having different characteristics and latencies (Fig 9 61) An early, large spike lasts about 1 msec , followed by a series of irregular discharges lasting for about 10 msec The first discharge can be evoked also by a quick stretch of an extensor muscle, so that it is the electrical counterpart of the stretch reflex The later, irregular discharge can be evoked alone by stimulation of a sensory nerve from skin, it is the electrical counterpart of a flexor reflex

FIBER TYPES

Afferent nerves are divided into four categories in accordance with their fiber diameter and velocity of conduction, as follows Group I,

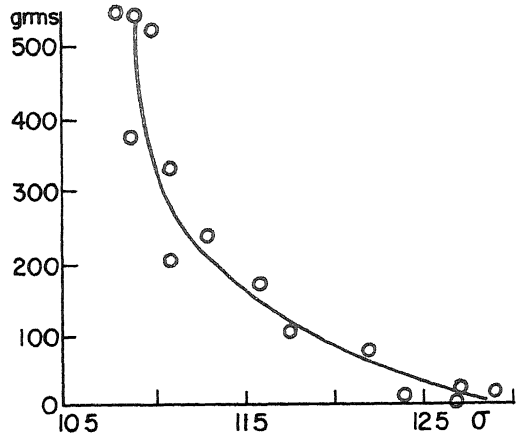


Fig 9 60 Relationship between force of reflex contraction and latency of the response As strength of a stimulus was increased force increased and latency became shorter (Creed et al 1932)

12 to 21 μm , Group II, 6 to 12 μm , Group III, 1 to 6 μm , Group IV consisting of C fibers less than 1 μm in diameter Groups I, II, and III_a correspond to fibers in the A category, group III corresponds to AΔ All somatic nerve trunks contain the afferent groups II, III, and IV Group I, containing the fastest afferent fibers is present only in nerves from muscles and is associated with afferents from the muscle spindle (group I_a) and the Golgi tendon organ (group I_b)

Motor nerves consist of two groups of fibers One group, the α motoneuron innervates the striated muscle fiber, is 9 to 13 μm in diameter and constitutes about 70% of all the effer-

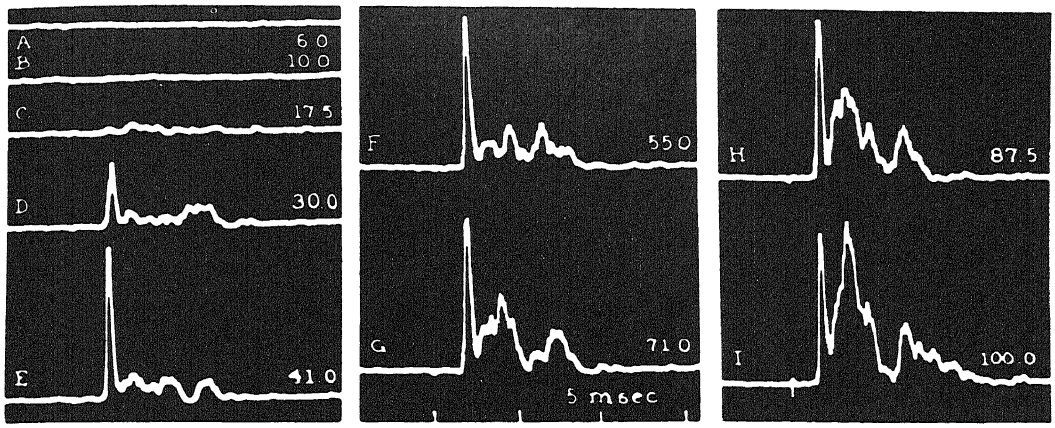


Fig 9 61 Dorsal root-ventral root reflex discharges from first sacral segment Records *D–F* show clearly the appearance and increase in the early spike of short latency as strength of the afferent stimulation was increased (figures on right of each record) The second later and more irregular waves are also shown (Lloyd 1943a)

ent fibers, the second group are the γ efferents, 3 to 6 μm in diameter and are the efferent fibers to the intrafusal muscles of the spindle

THE MONOSYNAPTIC REFLEX

A rapid reflex response of short latency is elicited by exciting type I sensory fibers This latency includes time for conduction through the fibers, plus time for excitation across any synapse As measured by stimulating a dorsal root and recording from the ventral root the shortest delay not accounted for by conduction in fibers is around 0 5 msec This appears to be the delay of a single synapse More precise measurement of the synaptic delay may be performed by stimulating neurons directly within the spinal cord in the region of the motoneurons (Fig 9 62) At low stimulus intensities, interneurons within the cord are fired, at higher stimulus strengths the motoneuron itself is excited The difference in time between the interneuronal and the motoneuron response is a measure of the synaptic delay By this direct method, the delay is found also to be 0 5 msec It therefore may be concluded that the early spike of the reflex discharge arises from a reflex arc in which only one synapse is present between the afferent and efferent neurons; no internuncial neurons are interposed between these. For this reason this reflex as recorded in this fashion is referred to as *monosynaptic reflex*

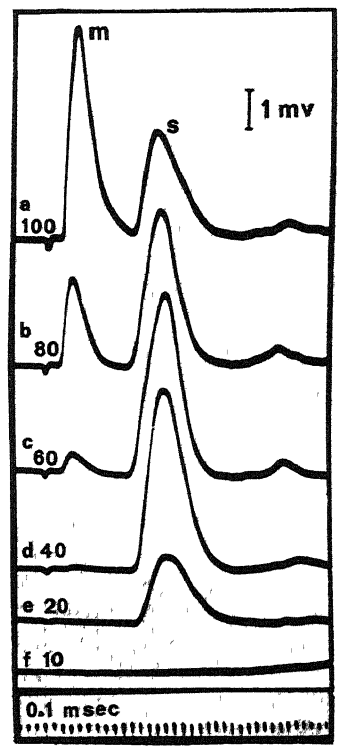


Fig 9 62 Measurement of central synaptic delay Records are from ventral root of cat after decerebration plus section of spinal cord Stimulus (slight downward blip in records) was applied by needle electrodes in intermediate grey matter of spinal cord Strength of stimulus was strongest for record *a* weakest for record *f* First spike represents direct stimulation of motor neurons second spike is result of activation via a single synapse Inverval between onset of the two spikes measures synaptic delay (Renshaw, 1940)

MULTISYNAPTIC REFLEXES

The electrical discharge which corresponds to the flexor reflex has a longer latency—of the

order of 3 to 5 msec It therefore must contain more than one synapse, and internuncial or connector neurons interposed between the input and output The afferent nerves involved in evoking the reflex are members of groups II, III, and IV Assuming that each synaptic delay is about 0 5 msec and that the transit time over the short interneuronal axons is negligible the number of neurons probably involved in the reflex is between 6 and 10 (reflex latency per synaptic delay) In this reflex action potentials continue to appear in the ventral root as an after-discharge, which results from persistent activity over delay pathways between the input and the final motoneuron At least two basic circuits exist which may give rise to an after-discharge (Fig 9 61) One arrangement of neurons consists of a series of delay pathways in parallel without any feedback from a subsequent to a preceding neuron This is called a *multiple chain network* A second neural configuration consists of a chain of neurons so arranged that some of the neurons further down in the chain may feedback to earlier neurons and re-excite them Such chains are called *reverberatory networks* for they give rise to activity which reverberates back to the first cells in the chain At present, the relative contribution of each network to after-discharge in multisynaptic reflexes is unknown

FACILITATION AND TEMPORAL SUMMATION

Spatial summation was attributed to the circumstance that two input pathways may contain nerve fibers which terminate on the same neurons The first or conditioning stimulus can be applied to the nerve from one head of the gastrocnemius muscle, the second or test stimulus which is strong enough to evoke a reflex is applied to the nerve from the other head of the muscle If the two stimuli are applied simultaneously, an augmentation of the response by as much as 300% may be observed If, however, the conditioning volley is separated from the test volley, it gradually loses its effectiveness so that when they are 10 msec apart the effect has almost disappeared It is clear from these experiments that a conditioning volley which causes no reflex response may give rise to facilitatory effects which may last for as long as 10 msec or more This kind of *residual facilitation* can occur in pathways where no internuncials seem to be present as in extensor nuclei of the cord, or the same results can be obtained in the flexor nuclei How a facilitation independent of internuncial delay networks can occur will be explained later in terms of the effects of the afferent volley upon the membrane potential of the motoneuron These experiments raise the question of whether temporal summation is possible in the nervous system, that is, whether two successive stimulations of a single afferent pathway can evoke a response greater than the predicted sum of the two individual responses The question is, does the residual facilitation (CES) last longer than the refractory period of the afferent pathway? The answer seems to be that temporal summation cannot occur naturally, but can be demonstrated when an electrical stimulus is applied directly to the motor neuron (Lorente de Nó, 1938a and b) In this case there is no afferent pathway with a measurable latency

ELECTRICAL RESPONSE OF THE MOTOR NEURON

The Excitatory Postsynaptic Potential (EPSP). As noted above, Eccles (1953) and co-workers inserted finely drawn-out microelectrodes made of glass and filled with KCl or NaCl solution into the cell body of motoneurons of the spinal cord for study of local potential changes They found that when the microelectrode is inserted, a resting potential of about −70 mv is observed, with the interior of the cell negative to the outside If the motoneuron is antidromically excited, i e , by stimulating its axon, the motoneuron is excited without the intervention of the synapse An action potential with an overshoot of about 20 mv appears Thus, antidromic excitation gives rise to an action potential in the postsynaptic cell, the motoneuron, which is no different in its general characteristics from that obtained from squid axon or the muscle fiber

The events of interest in synaptic excitation, however, occur when excitation is delivered orthodromically, i e , to the afferent neuron of the dorsal root, so that the presynaptic endings of this neuron are excited In such experiments in which a synapse is present, in recordings from the postsynaptic cell an additional potential appears which is a sign of synaptic depolarization Figure 9 63 shows the response of the motoneuron to a presynaptic volley The slow potential which appears just before the

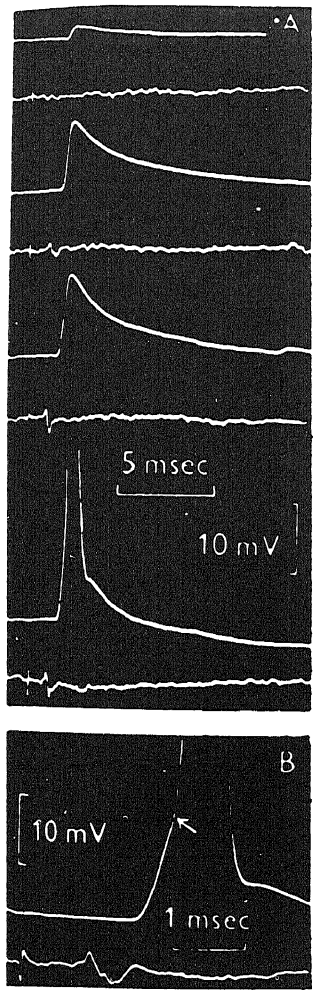


Fig 9 63 Intracellular recording of membrane potential changes in a flexor motor neuron as a result of stimuli of various intensities applied to a sensory nerve The upper record of each pair is the membrane potential the lower stimulus timing and strength Note increase in sweep speed in *B* to show the synaptic potential that anticipates the spike The first three potential traces show only the synaptic potential with no spike (Brock et al 1952)

action potential is called the excitatory postsynaptic potential or EPSP, and is the local, electrical sign of the CES of Sherrington It consists of a depolarization of the membrane which may last for 20 msec and can best be seen when the extent of depolarization is insufficient to give rise to an action potential When the membrane potential is decreased to a critical value of about −60 mv, a nerve impulse is generated and an action potential appears on the EPSP Temporal summation of such EPSP is also possible, if two subliminal volleys are sent in over the same nerve, each volley produces an effect which is manifested by an EPSP The EPSP's will then sum and if the critical level of depolarization is reached, an impulse will be set off

The EPSP is monophasic and nonpropagated, it represents a depolarization which is localized to the soma of the motoneuron It is not all-or-none in character since it can be augmented simply by increasing the intensity of the input volley Moreover, the EPSP's of different inputs can sum on a postsynaptic cell to produce a greater depolarization These characteristics show that the EPSP is produced in a process which is fundamentally different from that of the action potential, and illustrate the fact that EPSP has the properties Sherrington attributed to CES Several lines of evidence indicate that the EPSP arises from a depolarization of the postsynaptic membrane and an increase in permeability to both sodium and potassium ions

Inhibitory Potentials. If a monosynaptic reflex is evoked by stimulation of afferent fibers (orthodromic excitation), an inhibition of the reflex can be produced by stimulating an ipsilateral antagonist nerve as described more fully in the section on inhibition The effect of the inhibitory stimulus is to cause a hyperpolarization of the postsynaptic membrane in a direction opposite to that of the EPSP (Fig 9 64) This hyperpolarization consequent upon an inhibitory stimulus is called the inhibitory postsynaptic potential or IPSP It begins to develop a bit later than a monosynaptic EPSP, but its time course is otherwise the same as that of EPSP, of the order of 20 msec, and corresponds to the curve of inhibition obtained from reflex studies Indeed the IPSP is a mirror image of the EPSP in the same motoneuron The IPSP causes a hyperpolarization of the membrane and renders it less excitable according to the principles of peripheral neurophysiology Any EPSP which occurs during an IPSP will generate currents which will be less effective in causing excitation of the cell

The problem of determining the nature of the currents responsible for the IPSP has been attacked by using doubled-barrelled microelectrodes for simultaneously passing a current and recording the membrane potential Setting the membrane potential at −80 mv will cause the disappearance of the IPSP The potential −80 mv is the average of that determined by

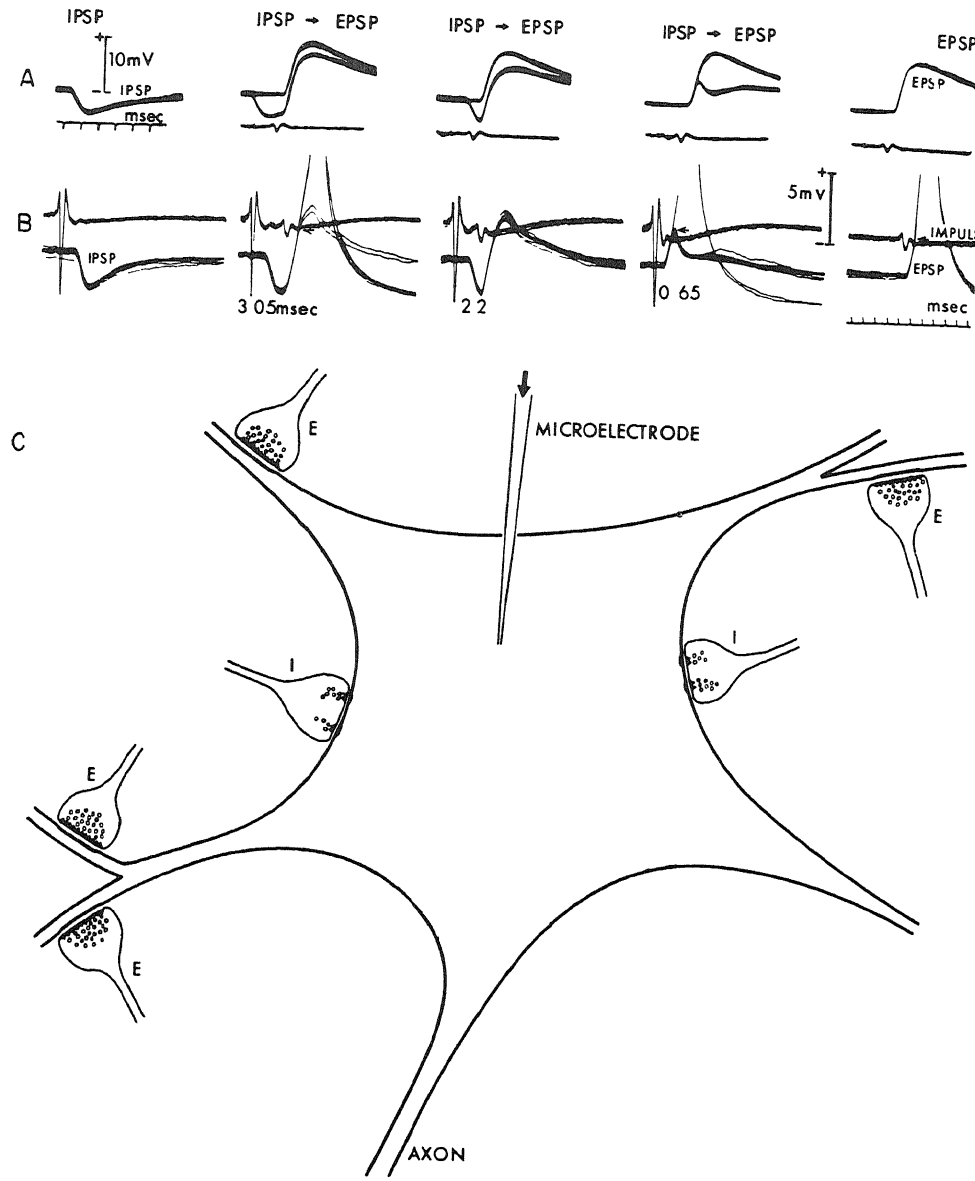


Fig 9 64 Diagram of neuron (*C*) with excitatory (*E*) and inhibitory (*I*) synapses and with two series of records of membrane potential recorded with microelectrodes (*A* and *B*) On these records the deflections that signal excitation or inhibition are labeled *EPSP* and *IPSP* respectively (Eccles 1967)

potassium (−70) and chloride (−90) suggesting that the IPSP results from an increased flow of K^+ out and Cl^- in, so that the inside becomes more negative Further evidence points to chloride as the principal ion involved in producing the EPSP The method of electrophoretic injection has been used to demonstrate this One barrel of a double microelectrode is filled with KCl and made negative so that this electrode effectively drives chloride ions into the cell, it is a hyperpolarizing current since it is internally negative The other barrel of the electrode records the IPSP Under these conditions the injection of chloride into the cell converts the IPSP elicited by a group I_a afferent volley from a hyperpolarization to a depolarization (EPSP) The injection of chloride had caused an increased internal concentration of chloride The arrival of the inhibitory stimulus then results in an increased permeability to chloride, but

because the chloride ionic gradient is now opposite to its normal direction, chloride now flows out and makes the inside of the cell less negative This causes depolarization instead of hyperpolarization A number of anions were injected internally and all those which were less than a certain hydrated ion diameter caused a reversal of the IPSP to an EPSP Apparently the inhibitory substance opens pores which allow ions below a certain size to pass Since chloride is the only anion which exists in a high enough concentration to flow in during such an inhibition, it is presumed that chloride is the main contributor to the inhibitory currents which flow

TYPES OF INHIBITION

Three kinds of inhibition have been extensively studied *a) postsynaptic, b) presynaptic,* and *c) Renshaw cell* or *recurrent* inhibition A fourth type, called indirect inhibition and extensively discussed by Sherrington, is now considered to be not a type of inhibition per se, but rather a form of occlusion

Sherrington and his associates (Fig 9 65) investigated the time course of the inhibitory process Using a flexion reflex as the indicator and a contralateral stimulus for inhibition, they found the central inhibitory process or CIS to be maximal when the inhibitory volley preceded the excitatory stimulation by about 80 msec

Postsynaptic or Direct Inhibition. Each of the three types of inhibition has been studied

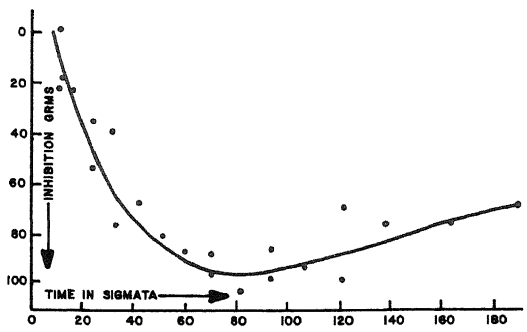


Fig 9 65 Latent interval for development of central inhibition Inhibition was created by stimulation of a contralateral cutaneous nerve at various intervals before ipsilateral stimulation evoked a flexion reflex Inhibition was maximal if inhibitory stimulus preceded excitatory stimulus by 80 msec Ordinate is percentage of inhibition, plotted inversely abscissa is duration of interval between the two types of stimuli (Creed et al 1932)

in similar experiments, and their time courses have been determined by utilizing two separate afferent stimuli that have opposing actions upon a given motor neuron pool For example, stimulation of I_a fibers in a dorsal root will initiate a monosynaptic reflex and EPSP on the motor neurons, whereas stimulation of type II fibers in an ipsilateral cutaneous nerve may inhibit the monosynaptic reflex by creating an IPSP on those same neurons The presence of the inhibitory process is detected by applying both stimuli in a single experiment, with variation of the interval between them Thus, if the inhibitory stimulus is given first by about 0 5 msec , there will be maximal inhibition of the subsequent monosynaptic response If the inhibitory stimulus follows the test stimulus by 1 msec , the inhibition arrives too late to block the reflex Utilizing this technique where the inhibitory stimulus is a "conditioning" stimulus (but not as Pavlov used the term), the time course of development and persistence of the inhibitory process can be revealed (Figs 9 66 and 9 67) It is this evidence that justifies the statement above that inhibition and IPSP have the same duration The observation that the inhibitory stimulus must precede the test stimulus by a minimal interval of 0 5 msec is taken to mean that the inhibitory pathway involves at least one more synapse than the excitatory reflex This interneuron in the inhibitory pathway possesses the ability to secrete a substance at its terminals which hyperpolarizes the membrane and leads to an IPSP Thus, a neuron which is normally excitatory to other neurons may exert an inhibitory action by the interposition of an inhibitory neuron Other mechanisms which have been proposed include terminating the neuron on a special region of the postsynaptic neuron so that it can have inhibitory effects on special regions of the postsynaptic cell

Postsynaptic inhibition is by definition the inhibition that has the characteristics that correlate with genesis and persistence of the IPSP as recorded using a microelectrode inserted into a motor neuron It is very susceptible to strychnine injection which can completely abolish it This effect of strychnine is the basis for the explanation of its convulsant activity— All postsynaptic inhibition having been removed, the slightest stimulus causes a tremendous extensor response

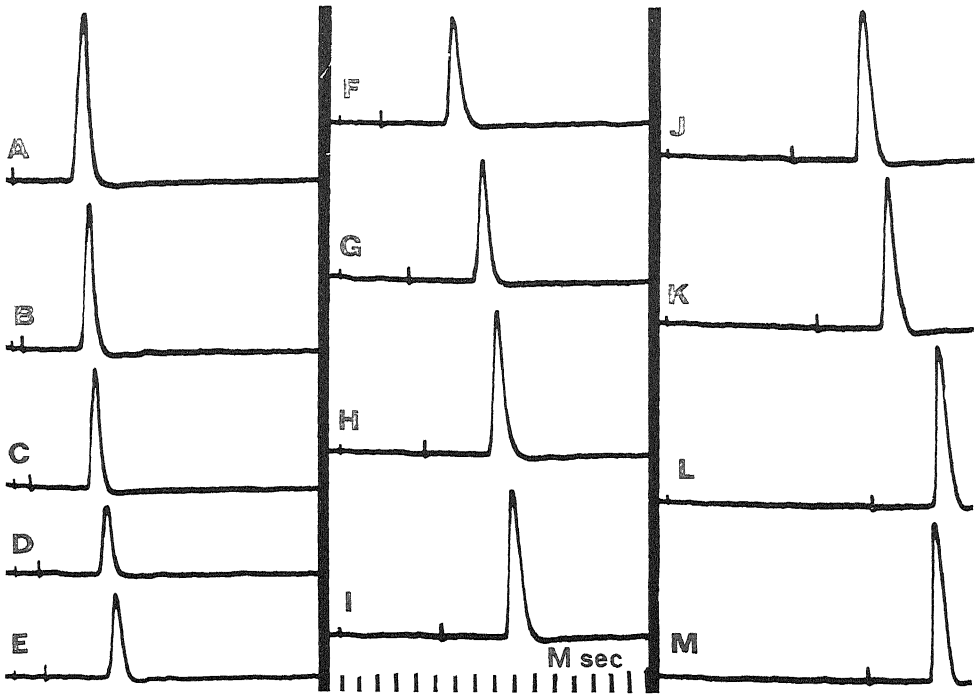


Fig 9 66 Inhibition of two neuron reflex to gastrocnemius induced by weak afferent volleys in deep peroneal nerve The height of the spike signals the strength of the reflex the interval between the two blips at the *left* of each record shows that there was variation in the time between the two stimulations As this interval was lengthened the strength of the response was at first inhibited then recovered its initial voltage (Lloyd 1946)

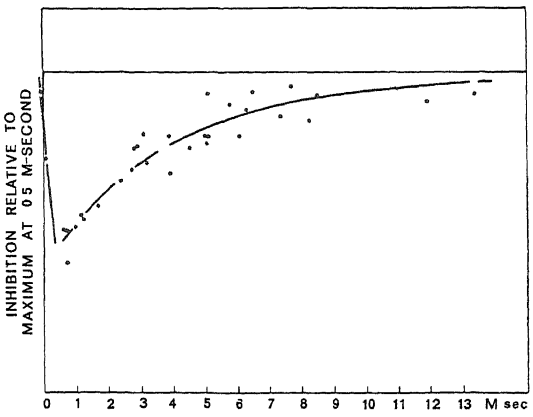


Fig 9 67 Strength of inhibited response relative to control plotted against interval between two stimuli The shortness of the latency for maximal inhibition cannot be fully accounted for Data from Lloyd 1946

Presynaptic Inhibition Early in the studies of the electrical responses evoked by reflex action, electrodes were placed on the surface of the spinal cord or on dorsal roots and large potentials which were positive to distant regions, called P waves were recorded It was suspected that this positivity might be the sign of some inhibitory presynaptic activity going on in the afferent terminals in the spinal cord To demonstrate presynaptic inhibition, a reflex response is elicited in extensor motoneurons This reflex will be inhibited if it is preceded by stimuli from any nerve entering the cord Group I_a and I_b fibers are most effective in causing the inhibition but groups II and III will also serve The time course of such an inhibition is shown in Figure 9 68 Unlike postsynaptic inhibition, maximum presynaptic inhibition is observed when the inhibitory stimulus precedes the excitatory by about 20 msec The inhibition may endure for as long as 200 to 300 msec and is extremely resistant to strychnine

Several lines of evidence indicate that this inhibition is presynaptic Recording within the motoneuron at the time that inhibition is produced shows no changes in the membrane potential, neither a hyperpolarization or a depolarization is produced by an inhibitory volley Instead, the magnitude of the EPSP which is set up by the reflex volley is diminished by the inhibitory stimulus and the change in magnitude of the EPSP parallels the curve of inhibi-

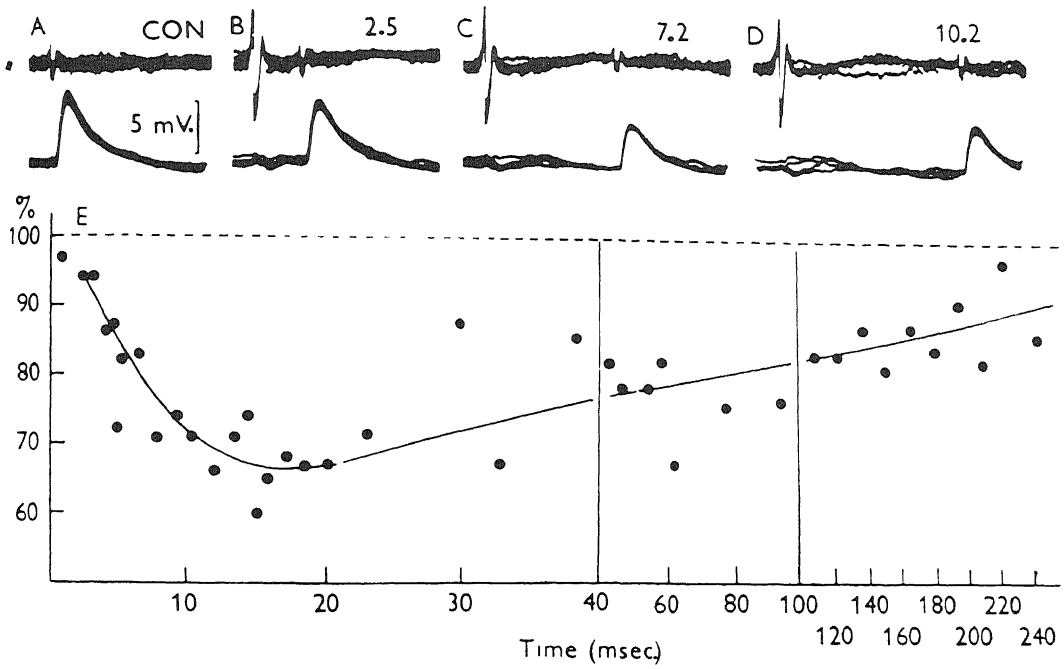


Fig 9 68 Presynaptic inhibition as a function of interval between inhibitory and the stimulating volleys *A–D* show time relations of volleys presynaptic inhibition is disclosed as variability in height of EPSP in lower of each pair of records

tion (Fig 9 68) The usual explanation for these observations is that the inhibition is due to some interaction at the terminals of the two groups of afferent fibers Activity in the inhibitory terminals apparently prevents activity in the excitatory endings A presynaptic depolarization may affect the amount of transmitter released from the presynaptic terminals Since long delays are involved in presynaptic inhibition, many interneurons are interposed between the first and last neurons of the chain

Renshaw Cell Inhibition. In 1941 Renshaw discovered that an electrical stimulus applied to motor axons and carried by them back into the ventral grey column causes an inhibition of all types of motoneurons at the segmental level This type of inhibition has therefore been called antidromic because the inhibition is evoked by "backwards" conduction via the motor roots into the spinal cord When such antidromic excitation is used, an after-discharge of quite high frequency is present in the recording from the ventral root Microelectrode recording has shown that the discharge does not originate in the motoneuron but in neighboring cells near the motoneuron—in cells which discharge with a high frequency when the antidromic excitation occurs At the same time the motoneuron is inhibited These neighboring cells are therefore believed to cause the inhibition The motoneuron displays a hyperpolarization which has all the characteristics of an inhibitory postsynaptic potential and the IPSP lasts for the period of time (Fig 9 69) corresponding to the discharge in the neighboring interneurons (known as Renshaw cells) An anatomical pathway has been suggested to explain these results (Fig 9 70) The motoneuron gives off a collateral to the Renshaw cells, the Renshaw cell axon returns to the motoneuron and inhibits it

The repetitive discharge of the Renshaw cell is presumed to be caused by an accumulation of acetylcholine at the junction between the motoneuron collateral fiber and the Renshaw cell It has been shown, for example, that the discharge of the Renshaw cell may be prolonged by anticholinesterases which allow acetylcholine to accumulate It is also possible to inhibit the discharge with β-dihydroerythroidine, a substance which blocks the action of acetylcholine The functional significance of this pathway is not clear It appears to serve as a general synaptic inhibitor and may act to limit the frequencies of impulses going to the motor end-plate

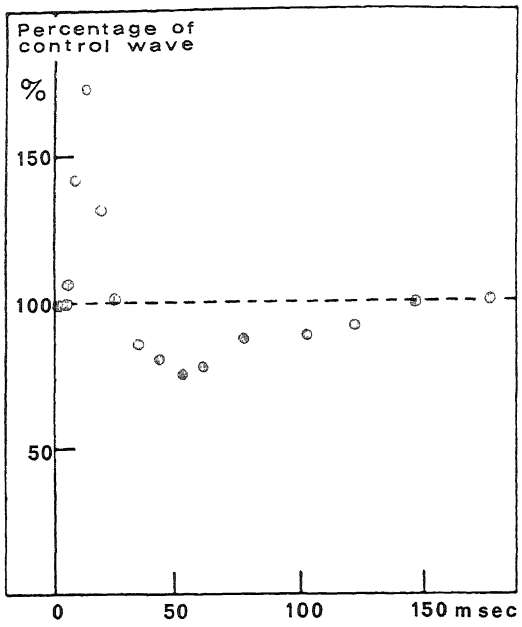


Fig 9 69 Time course for development of inhibition of the Renshaw type when antidromic stimulation of a motor neuron's axon precedes excitation via a dorsal root or sensory nerve (Renshaw 1940)

ELECTRICAL AND CHEMICAL TRANSMISSION

As noted earlier, events at a synapse may be interpreted according to either a chemical or electrical theory of synaptic transmission The chemical theory of transmission attributes the response of the postsynaptic cell to a chemical substance (transmitter) released from the presynaptic nerve terminals (Table 9 5) The electrical mechanism of transmission presupposes the activation of postsynaptic neurons by electrical currents flowing out from the presynaptic terminals into the postsynaptic cell Examples of each of these types of transmission may be found in various kinds of synapses among vertebrates and invertebrates It may be stated, however, that no evidence of electrical transmission at mammalian synapses has as yet been obtained

The chemical mechanism of transmitter release and action has been presented in section 1, chapter 3, and is summarized in Figure 9 48 γ-Aminobutyric acid is the compound most often considered as a possible inhibitory transmitter (Mitchell and Srinivasan, 1969, Curtis et al , 1970)

The Muscle Spindle and Stretch Reflexes

From the ultimate control process in skeletal muscle, (the release of calcium ions from the inner surface of the sarcolemma membrane), the origin of this control has been traced back along the course of the axons that originate in the spinal cord (Fig 9 71) In the cord the interaction of neurons has been considered, and the influence of two classes of afferent fibers has been described These are the I_a fibers from muscle that activate monosynaptic reflexes, and the types II, III, and IV fibers that provoke the flexion reflex Because the I_a fibers originate in muscle and are capable of evoking a stretch reflex, they function as the afferent component of the segmental control system for muscle It is necessary, therefore, to consider where they come from and what are the properties of the receptor or detector element they innervate (see Barker, 1962) It turns out that the detector element, itself, has a motor innervation different from that of the more typical muscle cells, and that this second or gamma (γ) innervation is a significant determinant of the sensitivity of the detector elements How the reflexes so established are integrated into attitudes of posture and into locomotion will be described in the succeeding chapter

When the position of the body is displaced backwards as in swaying, the extensor muscles of the leg such as the quadriceps are stretched As a consequence, the stretch reflex is called into play and the extensor muscles shorten to fix the joint The receptors involved in such a postural adjustment are the annulospiral endings of the muscle spindle, endings which are also responsible for the phasic and tonic components of the myotatic reflex

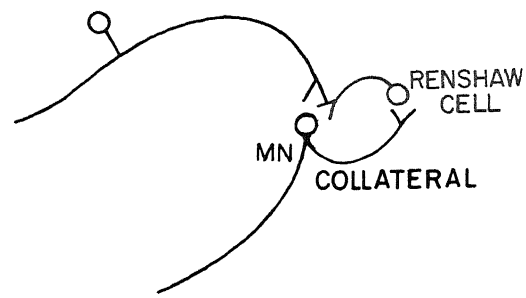


Fig 9 70 Diagram to illustrate recurrent or Renshaw type of inhibition *MN* is a motor neuron with a recurrent branch to the inhibitory or Renshaw cell

Table 9 5
Distribution of presumed chemical transmitters in the central nervous system

	Norepinephrine	Dopamine	Serotonin
	μg /gm	μg /gm	μg /gm
Cerebral cortex	0 03–0 24	0 00–0 10	0 04–0 24
Caudate nucleus	0 04–0 22	3 1–8 0	0 27–1 6
Corpus striatum	0 02–0 25	1 6–5 3	0 08–0 23
Olfactory bulb	0 05		0 11–0 35
Hippocampus	0 14–0 20	0 13	0 64
Septum	0 03	0 04	0 03–1 5
Hypothalamus	0 76–2 05	0 26–0 75	1 70–2 0
Diencephalon less hypothalamus	0 17–0 37	0 09–0 16	
Thalamus	0 13–0 24	0 01–0 07	0 22–0 24
Mesencephalon			1 0–1 7
Inferior and superior colliculi	0 11–0 16	0 10–0 13	
Substantia nigra	0 04	0 40	1–4
Red nucleus	0 22	0 19	
Brainstem reticular formation	0 35		
Cerebellum	0 06–0 17	0 03–0 02	0 27
Pons	0 04–0 41	0–0 11	0 19–0 70
Medulla	0 37–0 72	0 13	0 63–1 20
Area postrema	1 04		0 26
Cerebral white matter	0	0 42	0–0 13
Ant horn—spinal cord	0 18		
(Sup cervical ganglion)	6 8		

Data collated by Kety, 1967, p 495 in Quarton et al , 1967

TYPES OF ENDINGS AND FIBERS

A *muscle spindle* (Fig 9 72) contains both sensory and motor elements The motor parts are composed of two types of muscle fibers, intrafusal in location, having their nuclei centrally collected in the case of the nuclear "bag" fibers, or longitudinally distributed in the nuclear "chain" fibers Both are innervated by γ motor fibers, designated as γ_1 and γ_2, respectively, and believed to serve different functions (Matthews, 1964) The primary sensory role of the spindle is subserved by nerve endings within the nuclear bag These nerve fibers are afferent type I_s; they are larger than 12 μm in diameter and have conduction velocities of more than 90 m /sec Another, secondary group of afferents originates outside the nuclear bag in the myotube region, these endings are flower spray in type and their nerves are in group II (Fig 9 73) Their function is not clear, inasmuch as stimulation of this group of fibers apparently does not evoke a stretch reflex in extensor muscles Matthews has suggested that the primary endings are responsible for dynamic as well as static signals, that is, for detection of stretch and rate of stretch, as well as for length of the muscle The secondary endings are believed to be sensitive in a static sense to length

Muscle spindles are attached at each end either to a tendon, to other spindles, or to extrafusal muscle fibers They are said to lie "in parallel" with extrafusal fibers because shortening of the latter relieves tension on the spindles Conversely, the stretching of a muscle lengthens the extrafusal fibers and exerts tension on the spindles This may be an oversimplification, in that when the end of a spindle attaches to an extrafusal fiber the spindle acquires a "series" relationship to that fiber, at least These relationships are further compli-

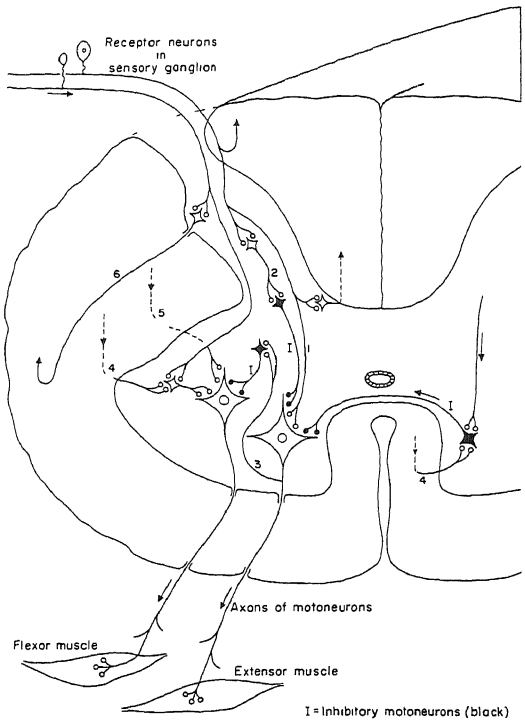


Fig 9 71 Diagram illustrating the interconnections that occur at a typical level of the spinal cord The flexor motor neuron (*left*) and extensor motor neuron (*right*) are intended to represent cells of two neuron pools Inhibitory neurons are in black and are labeled *1 1*) Connections of two-neuron arc extensor *2*) Multisynaptic pathway with inhibitory connector neuron *3*) Pathway of Renshaw type of inhibition *4*) Crossed and uncrossed input from other levels *5*) Corticospinal pathway in primates it comes directly from motor cortex *6*) Second order neuron of sensory pathway ascending to higher levels Bodian cited from Quarton et al 1967

cated by the activity of the intrafusal muscle fibers, to be described shortly

The *Golgi tendon organ* arises from the tendon of the muscle and is anatomically in series with the muscle When the muscle is stretched, the Golgi tendon organ is stretched and stimulated and when the muscle contracts, the Golgi tendon organ is likewise stretched and stimulated The Golgi organ gives rise to afferent fibers which are classified as I_b with a modal velocity of about 80 m /sec , somewhat slower than the I_a group The extent of stretch necessary to evoke a response from the Golgi tendon organ is greater than for the annulospiral ending

Gamma Efferent System The function of muscle spindles has become more difficult to describe with discovery of progressively more details of organization of the intrafusal mecha-

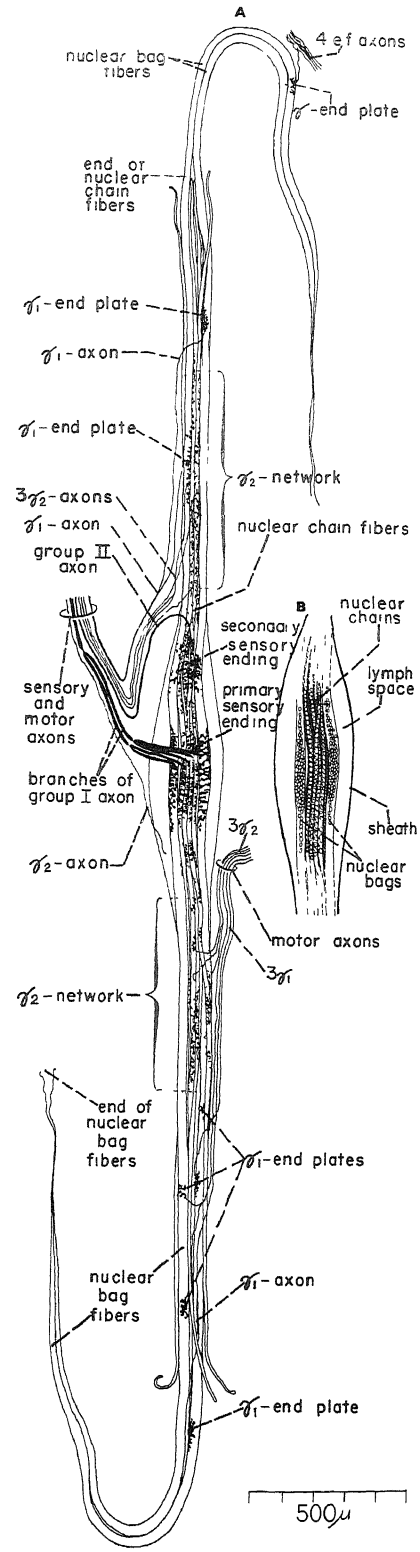


Fig 9 72 Drawing of structure and innervation of a muscle spindle (Boyd 1962)

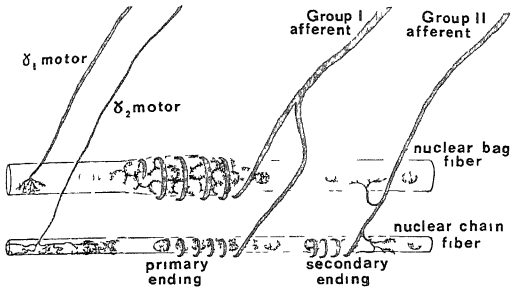


Fig 9 73 Diagram showing two types of ending primary and secondary on each of two types of fibers nuclear bag and nuclear chain Motor innervation is also shown (Matthews 1964)

nisms Not only is there the possibility that the primary ending is in parallel with most extrafusal fibers and in series with others, but as Fig 9 72 shows it is in parallel with the intrafusal fibers innervated by the γ_2 neuron, and in series with the element innervated by γ_1 These same relationships apply also to the secondary ending Consequently, it hardly seems adequate to describe the spindle as a detector that yields a negative feedback to the motor mechanisms of the spinal cord and merely inhibits any change in length of the muscle They do, of course, produce a negative feedback, but it has no simple relationship to the length of the muscle Perhaps the clearest way to explain the function of the intrafusal fibers of the mechanism is to assume that their contraction serves to control the sensitivity of the detector elements In doing this the intrafusal fibers partially isolate the sensory components from changes in the main or extrafusal fibers of the muscle Thus, contraction of a nuclear bag fiber may elongate the primary ending even though the muscle as a whole has shortened (Fig 9 74) With naturally occurring movement of the whole muscle this control of intrinsic tension permits the spindles to preserve sensitivity no matter whether the extrafusal fibers are short or long In this way the spindle may be able to signal change in the muscle under any condition of length or tension Nevertheless, the fact that relationships are in series as well as in parallel, and that intrinsic contractile elements have both types of relationship, further multiplies the possibilities of diversified function

The anatomical difference between the primary annulospiral and Golgi tendon organ is paralleled by a similar difference in their reflex responses to stretch When the annulospiral endings are in parallel with the extrafusal fibers, if a muscle containing such an ending is stretched, the annulospinal endings respond with an afferent volley in the I_a fibers, and a reflex myotatic discharge occurs in the extensor muscle These endings respond only to stretch and cease firing when the muscle shortens or relaxes On the other hand, the Golgi tendon organ responds to either contraction or extension of muscle However, it is necessary to stretch muscles considerably before any effective discharge occurs The reflex response in which the Golgi tendon organ is involved is called the *inverse myotatic reflex* or *clasp-knife reflex* If the extended limb of a decerebrate animal is flexed forcibly, considerable resistance is met until at a certain extension the limb suddenly gives way to the force The stretch of the extensor muscles beyond their normal position as a result of the forcible flexion activates the Golgi tendon organs and inhibits the extensor reflexes This type of inhibition is called *autogenic* since it arises from the very muscle from which the reflex itself was obtained

The properties of the muscle spindle that are significant for the maintenance of a fixed length of muscle are the following *a*) A passive stretch of the muscle and its spindles causes the primary endings to discharge and send impulses to the cord over group I_a fibers The discharge rate is high at the beginning of the stretch but falls off to reach a steady maintained, tonic rate of discharge from both primary and secondary endings The frequency of discharge of the spindle in the steady state is proportional to the extent to which the muscle has been stretched *b*) Adaptation of the spindle discharge is slight The tonic discharge

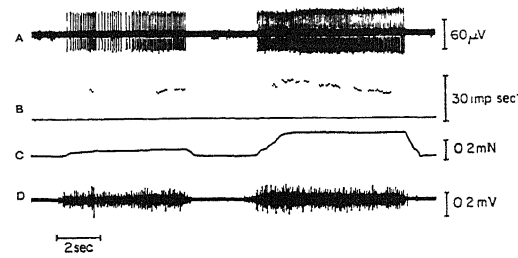


Fig 9 74 Discharge of sensory unit in muscle spindle (records *A* and *B*) as a result of tension in the intrinsic motor fibers accompanying voluntary contraction of the muscle Record *C* is force of muscle contraction *D* is electromyogram Data of A B Vallbo cited by Granit 1970 p 169

continues indefinitely as long as the muscle is kept stretched *c*) Contraction of the muscle (shortening of the muscle fibers) causes cessation of firing Shortening releases the spindle from the tension caused by extension, and the spindle is no longer mechanically excited This experimental finding demonstrates that the spindles are functionally in parallel with the extrafusal muscle fibers *d*) Contraction of the intrafusal fibers of the motor spindle results in an acceleration of the discharge from the spindle By stimulating the intrafusal fiber, the annulospiral endings are placed under stress and the endings are stimulated Contraction of the intrafusal fibers is equivalent to taking up the slack on the intrafusal fibers and placing the annulospiral endings under some degree of tension It is therefore the same effectively as passively stretching the muscle All these properties are important when the spindle acts as part of the reflex arc controlling the length and tension of a skeletal muscle

The γ efferent system presumably is involved also in the more rapid shortening of muscle such as occurs during voluntary movement Such movement can occur theoretically in either of two ways—a direct activation of the motoneuron via the corticospinal pathway, or excitation of the motoneuron indirectly through the γ efferent system (Granit, 1970) The γ efferents might be activated by descending fibers from the reticular formation Such stimulation would cause shortening of the intrafusal fibers, stretching of the annulospiral endings, and an immediate spindle discharge This discharge would then elicit a phasic reflex response of the extensor motoneuron Thus, a rapid voluntary movement might be evoked by intrafusal excitation This latter mechanism has been a subject for speculation for some years There is at least some evidence, however, that the γ system is not a necessary part of the voluntary control of skeletal muscle (Evarts, 1967)

THE KNEE AND ANKLE JERKS

A sharp tap upon the patellar tendon, when the knee joint is semiflexed, stretches the quadriceps extensor muscle and causes it to contract Similarly a blow upon the tendo Achillis causes a quick contraction of the calf muscles These brief contractions or jerks, as well as those described below for the upper limb, are "fractional examples of the stretch reflex" and are not due to stimulation of the tendon but to stretching of the muscle, a stretch of 0 05 mm or less in $\frac{1}{20}$ sec , sufficient to elicit the response A voluntary action, such as clasping the hands together, reinforces the reflex and increases the force of the jerk The spinal center for the knee jerk in the human subject lies in the 2nd, 3rd, and 4th lumbar segments of the cord, it is innervated through the anterior crural nerve The center for the ankle jerk is situated in the 1st sacral segment, the peripheral nerve is the sciatic

Any condition, such as decerebration or injury to the descending motor pathways, which enhances the stretch reflexes increases the tendon jerks [1] They are abolished by an injury or disease involving the efferent or afferent limb of the reflex arc or the center itself (anterior horn cells) On account of its brief twitchlike character the knee jerk is sometimes spoken of as the "phasic reaction" of the stretch reflex The knee jerk is less affected by abnormal states, e g , spinal transection, anesthetics, circulatory failures, etc , than are the postural reactions In an animal such as the dog or rabbit, the knee jerk returns in a few minutes after spinal transection, in the monkey only after some days In man, though postural reflex activity is entirely lost and the muscles are quite flaccid, the knee jerk is elicitable in some degree within 3 weeks after a complete transverse lesion of the spinal cord

The smooth and steady character of the contractions brought about through the stretch reflexes concerned with the maintenance of posture is due to the asynchronous sensory impulses set up by the numerous stretch receptors and the resulting asynchronous reflex discharge of motor impulses to the muscles The muscle fibers are, therefore, never all re-

[1] When the electrical responses are recorded from the muscle of an animal during the elicitation of a tendon jerk, cessation of the action currents is found to occur during the actual contraction of the muscle This, the so-called "*silent period*," is probably due to the muscle spindles being relieved of stretch as the muscle shortens An additional factor is the synchronous nature of the discharge down the motoneurons, for the latter, being all excited in unison, pass also simultaneously into the subnormal phase

laxed nor all contracted at the same time The sharp contraction of the quadriceps characteristic of the knee jerk, on the contrary, is brought about by the discharge of a volley of synchronous impulses which causes the fibers of the muscle to contract in unison

Clonus Under certain conditions a muscle or group of muscles instead of contracting smoothly and continuously may do so rhythmically in a series of rapidly repeated movements or jerks Thus, in lesions of the central nervous system associated with hypertonus, sharp passive dorsiflexion at the ankle joint with maintenance of the foot in the dorsiflexed position by light pressure upon the sole, evokes a stretch reflex which consists of a series of rhythmical contractions of the stretched muscles (ankle clonus) The sudden dorsiflexion causes a sufficient number of stretch receptors to discharge a synchronous volley of impulses along the afferent nerve to the spinal centers and a synchronous efferent volley to be discharged to the muscles The continued stretch of the muscle caused by keeping the ankle in the dorsiflexed position sets up another synchronous volley as soon as the motoneurons have recovered each time from the subnormal phase of the preceding one Thus, the clonic movements continue so long as the muscles are kept upon the stretch Clonus can be evoked in a similar fashion in other situations It seems convenient to regard all such rhythmic activity as arising from feedback loops that oscillate because their normal damping mechanisms are not working properly (see also p 9-113)

SOME OTHER STRETCH REFLEXES (DEEP OR TENDON REFLEXES) OF CLINICAL IMPORTANCE

The *jaw jerk* Tapping the chin with the mouth partly open, and jaw supported, stretches the masseters which contract and jerk the jaw The reflex center is in the pons

The *biceps jerk* is elicited by a sharp tap upon the biceps tendon in front of the elbow joint, the response consists of a quick contraction of the biceps with flexion of the elbow The center for the reflex is situated in the 5th and 6th cervical segments of the cord, it is innervated through the musculocutaneous nerve

The *triceps jerk* is evoked by a blow upon the triceps muscle just above the olecranon process of the ulna, contraction of the muscle and extension of the elbow result The center for the response lies in the 6th and 7th cervical segments, the peripheral nerve is the musculospiral (radial)

The *supinator jerk* consists of contraction of the supinator muscle and flexion of the elbow, it follows a blow upon the styloid process of the radius The center lies in the 5th and 6th cervical segments of the cord, the peripheral nerve is the musculospiral

Rossolimo's reflex, which is seen with hyperactivity of the stretch reflexes in the lower extremity, consists of flexion of the toes, including the hallux, when the toes are flicked on their plantar surfaces A similar response (*Hoffmann's reflex*) which consists of a sudden flexion of the terminal phalanx of the thumb can be evoked by flicking the finger tips

5 / SUBCORTICAL MECHANISMS CONTROLLING POSTURE AND LOCOMOTION

CORTICAL VS SUBCORTICAL MECHANISMS

Neural mechanisms controlling movements of the body are extremely complex and as yet incompletely understood "Voluntary" movements[1] are dependent in primates (monkeys, apes, and man) upon the precentral motor cortex The same is not true for lower animals *Birds* and *reptiles* have practically no cerebral cortex (Ariens Kappers, 1929), and their well-organized movements are controlled by nuclear masses more or less comparable to the basal ganglia of the primate brain In higher animal forms such as *rodents* the cerebral cortex is present but of relatively little importance so far as muscular activity is concerned In the *carnivores*, e g , cat and dog, a specialized portion of the cerebral cortex that lies quite far forward (the sigmoid gyrus) functions as a motor cortex If it is stimulated electrically, movements can be produced in the contralateral extremities, and if the sigmoid gyrus is removed a transitory contralateral paralysis ensues However, the cerebral cortex is not essential for useful movement of the skeletal musculature It can be removed from both sides with only a temporary paralysis, the animal will recover the ability to sit, stand, walk, or run (Bard and Macht, 1958) In *primates* the situation is quite different Here the cerebral cortex is much more important even though there exists a considerable difference between the various animals If one removes the precentral motor cortex, i e , the gyrus ahead of the central sulcus, or even the entire cortex from one cerebral hemisphere in a monkey, a complete paralysis of the opposite side of the face and of the opposite arm and leg will result Useful movement will soon begin to return and in a few weeks the animal will be running about climbing the walls of his cage and hopping about on his trapeze in an agile fashion, although a partial paralysis of the involved arm and leg will be obvious The ape, too, will recover, though not so completely as the monkey, whereas removal of the precentral motor cortex in *man* will result in a contralateral paralysis which, though it will improve, will always be severely disabling After such an operation a man will become able to walk, although with a limping gait dragging and circumducting his leg He will never recover useful movement in his upper extremity and will probably have no movement at all in his hand and fingers

The recovery in primates following removal of the precentral motor cortex is due almost entirely to ipsilateral innervation from the intact cerebral hemisphere, for if the remaining precentral motor cortex is removed a severe and enduring paralysis *in all four extremities* is the immediate result It is true that if the cortex is removed piecemeal in several operations over months of time and if the animal is given very special care occasionally it is possible to produce a bilaterally decorticate monkey that is able to make a few awkward but very limited voluntary movements At best such an animal is almost completely disabled In most instances he is totally paralyzed except for a few useless sterotyped grasping movements In view of the demonstrably greater importance of the precentral motor cortex in man as compared with the motor cortex of cats it is obvious that in the human being the production of useful voluntary movements is dependent upon this important area of the cerebral cortex It is likewise evident that the subcortical centers (basal ganglia, red nucleus, substantia nigra, reticular formation, cerebellum, and spinal cord) are important in

[1] A "voluntary" movement is one that I believe I could make if I wanted to

the production of well-coordinated, useful muscular activity, but in man these centers cannot function adequately in the absence of the precentral motor cortex

CORTICAL PROJECTIONS TO LOWER MECHANISMS

The precentral motor cortex gives rise to corticospinal fibers which arise from the precentral gyrus and descend to the anterior gray horns of the spinal cord (Fig 9 75) These fibers are concerned with voluntary muscular activity and primarily with delicate, precise, well-coordinated movements of the distal parts of the extremities, particularly the digits The relationship of the corticospinal fibers which arise in the parietal cortex to voluntary muscular activity is unknown It is certain, however, that this parietal component of the corticospinal system is not capable of producing useful coordinated movements independent of the precentral motor cortex

The precentral motor cortex gives rise also to many other descending fibers with motor functions There are fibers to the basal ganglia, thalamus, red nucleus, substantia nigra, the pontine nuclei, the inferior olivary complex and the reticular formation These, too, are concerned with the production of voluntary muscular activity In the absence of the corticospinal tract such movement is apt to be cruder and less delicate but still useful Exactly which of these many fiber systems—corticocaudate, corticopallidal, corticonigral, corticopontine, etc —is concerned in the production of voluntary muscular activity is unknown

These neural mechanisms are also concerned with other functions of great importance in the organization of muscular activity These include the control by inhibition and facilitation of *1)* the various postural reflexes, *2)* of various automatic and associated movements, such as swinging of the arms when walking, and *3)* of various visceral and vasomotor activities such as respiratory movements, blood pressure, pulse rate, the vascular bed, gastrointestinal motility, etc , that accompany movements of other parts of the body

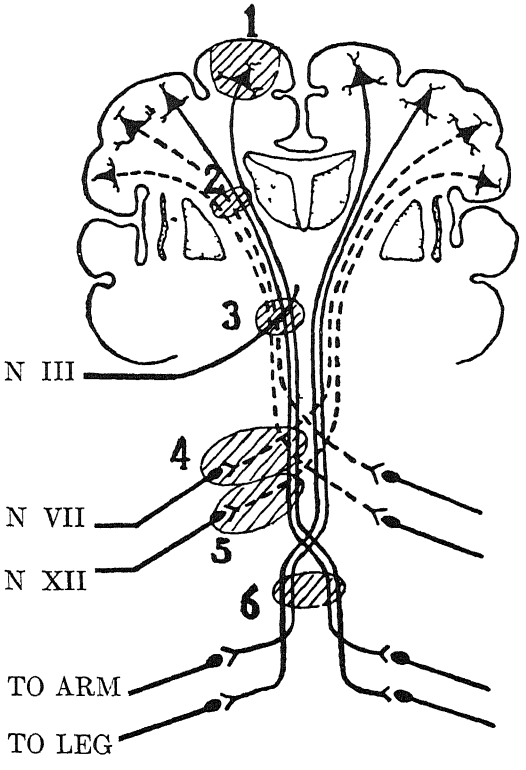


Fig 9 75 Diagram illustrating the clinical signs that follow typical lesions of the central nervous system *1)* Lesion of cerebral cortex causes monoplegia or hemiplegia depending upon its extent *2)* Lesion in internal capsule may cause hemiplegia *3)* Lesion of midbrain may involve descending motor fibers to opposite side of body and an ipsilateral eye muscle nucleus *4)* or *5)* Lesion in pons or medulla has results similar to location 3 but involving other cranial nerves *6)* Transection of spinal cord causes paraplegia

MECHANISMS OF POSTURE

Tone of Skeletal Muscle

As a result chiefly of the work of the Sherrington school the word "tone" or "tonus" as applied to skeletal muscle acquired a specific, technical meaning Muscle tone is the steady reflex contraction of the muscles concerned in maintaining the posture characteristic of a given animal species Tonus has its basis in the stretch reflexes, and in a standing animal its seat is therefore mainly in the *antigravity muscles* In most mammals these are extensor muscles, the muscles that show an exaggeration of tone following decerebration

In man the antigravity muscles and consequently those which exhibit the greatest degree of tone are the retractors of the neck, the elevators of the jaw (masseters), the supraspinatus, the extensors of the back, the ventral muscles of the abdominal wall (probably), and the extensors of the knee and ankle (vastocru-

reus, gastrocnemius, and soleus) When these muscles are completely relaxed, as in an unconscious person, the body collapses In the healthy conscious person, stretch reflexes are largely instrumental in preventing this occurrence

Though the fundamental basis of tonus in voluntary muscle is the myotatic reflex centered in the spinal cord, the tonic state is influenced profoundly by higher centers Impulses from labyrinthine and neck muscle receptors exert an influence upon a background of tonus established through lower spinal centers Similarly, pathways from cerebellar, midbrain, and cerebral neurons convey impulses which, impinging upon the final common path, are capable of altering the degree of tonus, or effecting finer adjustments in the tonic state and of maintaining its normal distribution between groups of muscles (Fig 9 82, p 9-108) The tone of a given group of muscles may also be influenced through the spinal centers by impulses arising in other muscle groups (e g , neck muscles and the muscles of the digits, ankle, and wrist, as in the positive supporting reaction) and in skin receptors

When the spinal cord is separated from higher centers there is an a-reflexic state of muscular flaccidity Gradually this gives way to a state of hyperreflexia The extensor tone, however, never is as pronounced as in decerebrate rigidity, whereas the flexor reflexes are more active

A feature of tonic contraction is its economy in the expenditure of energy Posture is maintained for long periods with little or no evidence of fatigue Examples are found in decerebrate rigidity, in the maintained closure of the jaws, standing or sitting in the normal person, and in the clasping reflex of the frog The economy of energy is possible because different groups of muscle fibers contract in relays, only a proportion of the total number of fiber groups of the muscle being active at any moment Thus, active fiber groups mingled with inactive groups are scattered throughout the muscle The alternating periods of rest and activity of the muscle groups explains the ability of the tonic contraction to be maintained for so long without showing fatigue

RED AND WHITE MUSCLE

The skeletal muscles of many animals, e g , birds, rabbit, cat, etc , can be clearly divided into two types, *a) red or dark,* and *b) white or pale* (Creed et al , 1932) Of the two, the red muscles are composed of smaller fibers having a granular and more opaque appearance, possessing more distinct longitudinal striations but less pronounced cross striations, and containing a larger proportion of sarcoplasm These fibers contract more slowly, fatigue less readily, and are tentanized at a slower rate of stimulation than are the pale fibers The white or *pale* fibers are translucent, show prominent cross striations and a small quantity of sarcoplasm Those muscles which execute rapid movements are usually, though not invariably, of the pale variety, whereas the slower movements are carried out chiefly by the dark muscles It is probable that all muscles are a mixture of the two types of fiber, but that in some the red, in others the pale type, predominates

Experimental Preparations for Study of Posture

THE SPINAL STATE

Transection of the cord produces an immediate flaccid paralysis of the muscles below the point of section Immediately after section in the lower cervical region the limbs hang limply, the muscles being quite toneless The stretch reflexes and other extensor responses cannot be elicited and the knee jerk is abolished The blood pressure falls and vascular and visceral reflexes are unobtainable This condition is called *spinal shock* Its duration varies from species to species The higher the position of the animal in the phylogenetic scale the more profound is the shock and the slower is the recovery In the frog the duration of spinal shock is brief, a few seconds In the rabbit, the knee jerk returns within a few minutes, in the cat within an hour, and in the dog somewhat more slowly In the monkey it does not reappear for several days Other extensor reflexes remain in abeyance for a much longer time In the cat and dog the picture immediately following spinal transection is different from that seen after decerebration The extensor stretch reflexes are less active while the flexor responses to nociceptive stimuli are more easily elicited Consequently, if the spinal cord of a decerebrate preparation is sectioned, the exaggerated extensor tone is replaced (below the section) by an imbalance in the sense of increased flexor tone

Spinal shock also follows section through the

medulla below the vestibular nuclei (*decapitate preparation*)

In the cat and dog recovery from spinal shock gradually takes place in a few days The blood pressure is restored to normal and the vascular reflexes can again be obtained The reactions of the extensor muscles return, and the animal is able, though imperfectly and very briefly to support the weight of the body when placed upon its feet (They cannot "stand" as does a normal animal, of course, because spinal motor mechanisms are cut off from those in the brainstem that are required for postural adjustments See below)

Cats, dogs, and rats that have recovered from spinal shock exhibit neural reactions that provide information about the function of spinal mechanisms For example, they eat normally, and digestion is not impaired Defecation is at first seemingly uncontrolled, with diarrhea and/or constipation, but later acquires a reflex character with stools that are formed normally Urination has a similar course, with first of all a retention of urine in the bladder that can be relieved only by artificial means such as catheterization or manual expression After a few days, however, the bladder empties periodically on a reflex basis

Spinal animals also exhibit reflex performance that is surprisingly complex Two notable examples are the rhythmic movements of the scratch reflex—an intersegmental reflex in which the hind leg moves rhythmically in response to stimulation of the skin over the shoulder—and stepping or the "mark time" reflex, in which the two hind limbs alternately extend and flex in movements like those that produce locomotion

Explanation of Spinal Shock. Spinal shock is attributed to the removal of impulses which in the intact animal descend from higher centers to reinforce the spinal centers That it is due to this and not simply to an inhibitory effect of the local injury itself seems clear from the fact that after an animal has recovered from spinal shock, a second transection made behind (lower than the original one) does not cause a return of the shock state In *the cat and dog* the flexor reflexes are evidently dependent only to a minor extent upon the higher centers, since they are capable of being executed by the spinal centers alone They are also seen in their most exaggerated form in animals or human beings with a complete transection of the spinal cord The fact that the extensor stretch reflexes are less active in the spinal animal is due to the loss of facilitatory impulses from the reticular formation and the vestibular nuclei Recovery of reflex activity following transection of the spinal cord is in considerable measure dependent upon the condition of the animal or patient Infection or impaired nutrition greatly retard recovery

DECEREBRATE RIGIDITY

This is the term applied to the sustained contraction of the extensor muscles which supervenes upon transection of the brain stem at any level between the superior colliculi and the vestibular nuclei [2] It was first studied and described in detail by Sherrington (1906) The animal assumes a characteristic attitude with limbs stiffly extended, head retracted, jaws closed, and tail horizontal or erect When placed upon its feet the limbs support the weight of the body The position is a *caricature* of the normal standing position The knee jerk and other stretch reflexes are exaggerated The righting reflexes are abolished, tonic neck and labyrinthine reflexes are retained

In some instances there is increased reflex excitability of the flexor muscles as well, but the characteristic feature of decerebrate rigidity is the tonic contraction of the muscles which maintain the posture of the body against gravity, the *antigravity muscles* It may be pointed out that in the frog whose natural posture is squatting with flexed thighs, legs, and arms, it is the flexor muscles which are the site of decerebrate rigidity Also in the sloth, whose habit it is to remain for long periods suspended from a tree branch, flexor rigidity is the characteristic result of decerebration In the pigeon the flexor muscles which maintain the resting (folded) position of the wing exhibit rigidity, and in the ape the muscles which hold the elbow semiflexed when the body is erect show increased tone In man abolition of the function of the cerebral motor cortex results in a state of increased tone in the flexors of the arms and in the extensors in the lower extremities, whereas decerebration at the mesencephalic level is associated with extensor spasticity in all four extremities

Tonic Labyrinthine and Neck Reflexes in

[2] Decerebration can also be produced by tying the common carotids and the basilar artery at the center of the pons and thereby depriving the forebrain of its blood supply

the Decerebrate Preparation. Labyrinthine proprioceptors are responsible for the tone of the neck extensors, consequently, after destruction of the labyrinths in the decerebrate preparation the head is no longer held erect but falls into the fully flexed position (Fig 9 76) The rigidity in the forelimbs in turn is maintained through proprioceptive reflexes initiated in the extended neck muscles as well as directly through labyrinthine reflexes (Fig 9 77) After labyrinthectomy, therefore, the flexed position assumed by the head sets up proprioceptive impulses from the neck muscles which lead to a reduction in the extensor tone of the forelimbs, these then become strongly flexed upon the chest If the neck muscles of the labrinthectomized animal are deafferented, movements of the head are without effect upon the extensor muscles of either fore or hind limbs

Lengthening and Shortening Reactions. Attempts at passive flexion of the extended limb of the decerebrate (or chronic spinal) preparation are met by considerable resistance If the force is increased beyond a certain point the stretch reflex (upon which the resistance depends) is inhibited and the limb gives way suddenly (*clasp-knife effect*) It may then be easily flexed to any degree and remains in the new position The elongation of the extensor muscles which permits the flexion of the limb is called the *lengthening reaction* If the limb is then moved again into the extended position the extensor muscle shortens adaptively, and again resists the limb being bent This is the *shortening reaction* These reactions give the muscles a certain plastic quality The giving way of the extensor contraction in the lengthening reaction is ascribed to inhibitory impulses from the Golgi tendon organs that are set up when the stretch stimulus reaches a certain intensity With inhibition of the forcibly flexed knee extensor, the knee extensor of the opposite leg is excited, this is known as *Phillipson's reflex*, a special instance of the principle of reciprocal innervation

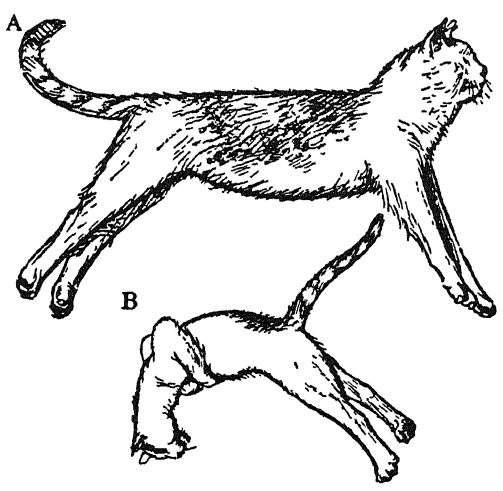


Fig 9 76 *A* Posture of cat following midcollicular decerebration there is overactivity of all the antigravity muscles *B* Decerebrate cat following labyrinthectomy the head is no longer supported against gravity and flexion of the neck leads to reflex inhibition of extensors of fore limbs

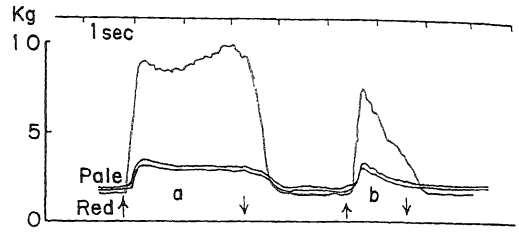


Fig 9 77 Influence of neck reflexes on stretch reflex in decerebrate preparation With triceps muscle stretched to induce reflex in both heads (red muscle *upper line* pale muscle *lower tracing*) the force of contraction is increased when neck is dorsiflexed and decreased with ventriflexion (*arrows*) (Creed et al 1932)

Mechanisms Responsible for Decerebrate Rigidity Decerebrate rigidity is quite evidently a "release" phenomenon, brought about by the removal of inhibitory mechanisms that lie ahead of the plane of section and that normally inhibit the neurons that now cause the rigidity (This is the inverse of the concept that spinal shock is a result of loss of rostral facilitatory mechanisms) The level of the neurons that induce the extensor rigidity can be specified, in cats it is the level of the pontile and upper medullary tegmentum If the plane of decerebration is at or above the rostral limit of the mesencephalon, little or no rigidity occurs, this indicates that the inhibitory mechanism lies in the rostral mesencephalic tegmentum Consequently, if the first section is made superior to the mesencephalon, and then after some minutes sections are made at progressively lower levels, the animal is at first not rigid but becomes so when the rostral mesencephalon is removed With the lower cuts the rigidity may increase to a maximum, and

then decrease again finally to disappear when the plane of section is at the lower border of the pons As noted in section 6, and in chapter 7 of this section, typical changes in pattern of respiration accompany these changes in posture

In a cat made decerebrate by a midcollicular section, the cerebellum has a curious relationship to the rigid state When an electrical stimulus is applied to the cortex of the anterior lobe, the rigidity is decreased, if this portion of the cerebellum is ablated, the rigidity increases For this particular mechanism, at least, this part of the cerebellum seems to have an inhibitory function analogous to that of the rostral mesencephalic tegmentum

Granit (1955) has shown that the brainstem mechanism that produces the rigidity does so by way of the stretch reflexes of antigravity muscles It facilitates the strength of the intrinsic control of muscle stretch receptors by activating the γ motor fibers This makes the receptors overly sensitive to stretch, intensifies their stretch reflexes, and leads to the decerebrate state The factor that determines that it is the extensor and not the flexor muscles that become hyperactive must be assumed to be the way the reticular facilitatory mechanism is organized Both extensor and flexor muscles have the same kind of control system from the level of the medulla down through spinal cord and out to the muscles The fact that removal of higher levels leads to only extensor rigidity, however, indicates that it is the extensor components of the reticular-spinal system that are usually subject to powerful inhibition from above The components that are so inhibited seem to include the reticular formation at the pontile level, the vestibular nuclei, and the central nuclei of the cerebellum These latter nuclei, in fact, can be regarded functionally as displaced components of the bulbar facilitatory system

From the outline given in chapter 2, it is apparent that there is no lack of descending tracts that might impose a tonic inhibition upon these tegmental and cerebellar nuclei The corticobulbar and corticopontile fibers, however, probably do not need to be considered in this connection, inasmuch as removal of the cerebral cortex of a cat leads to little evidence of extensor rigidity Furthermore, the lack of rigidity following high mesencephalic section also argues against the significance of telencephalic structures The most plausable conclusion, for the cat at least, is that both the inhibitory and excitatory mechanisms for the rigidity reside in the brainstem tegmentum and cerebellum The principal collection of cells found at the level of the critical plane of section is the red nucleus (Fig 9 78) When it retains its connections with brainstem and cerebellum, there is little rigidity, when the plane of section is below this nucleus, ridigity is marked If the inhibition normally exerted upon extensor reflexes does, in fact, arise in the complex of systems that includes the inferior olive, the Purkinje cells and nuclei of the cerebellum, and the red nucleus, this might explain why maximal rigidity follows concurrent removal of red nucleus and anterior cerebellar cortex

One other significant feature of the decerebrate state should be noted, namely, the evidence it offers for opposing facilitatory and inhibitory mechanisms in the brainstem (Fig 9 79) That such opposed mechanisms are almost ubiquitous in control systems will become evident as they are more fully described In general, the two types of components may lie together at any level of the neuroaxis, but occasionally they are at different levels as in the case of decerebrate rigidity, and also the expiratory and inspiratory mechanisms of the medullary respiratory system Longitudinally, these systems are organized so that there is increasing complexity of function at progressively higher levels of the nervous system

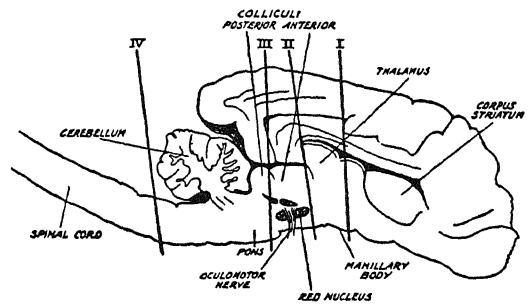


Fig 9 78 Levels of section through the brain of cat to illustrate typical results *I) Thalamic animal* Righting tonic labyrinthine and neck reflexes are retained little change in muscle tone *II) High decerebration* Minimal decerebrate rigidity *III) Midcollicular decerebration* Extensor rigidity may be maximal If vagi are cut, animal may show apneusis *IV) Medullary or decapitate animal* Condition depends completely upon level of section Usually there is no rigidity respiration may be gasping in type

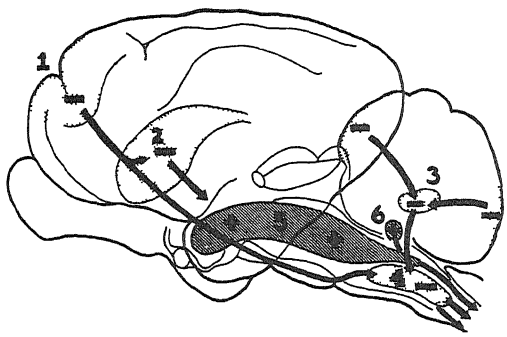


Fig 9 79 Outline drawing of structures in cat s brainstem in cerebellum and cerebral cortex Stimulation of regions marked + facilitates spinal reflex activity whereas regions marked – inhibit reflex responses (Lindsley et al 1950)

Thus, at spinal levels they control through opposing actions the strength of reflex responses, at the medullary level they evoke inspiration or expiration, raise or lower the blood pressure, accelerate or slow the heart, and strengthen or weaken the tone of antigravity muscles In the hypothalamus they raise or lower body temperature, and increase or decrease water and food intake Of these examples, however, posture is in many respects the one easiest to analyze, and how it is controlled may be used as a guide in study of the other systems

"THALAMIC ANIMAL

It is relatively easy to perform an operation in which most of the cerebral cortex is removed on both sides, with most of the white matter beneath it and accessible portions of the basal ganglia What remains is the diencephalon, together with variable amounts of the basal ganglia that were left behind The resulting preparation has been called a "thalamic" animal This name is misleading, because it is inconceivable that the thalamus under these conditions could have anything approaching its normal function Those thalamic nucleic most directly related to cerebral cortex must be seriously affected by the operation, and neurons sending their processes to the cortex can scarcely retain normal function after their processes are cut Consequently, no one knows just which subcortical structures are capable of activity, nor at what level in the neuroaxis the behavior exhibited by these animals may originate Nevertheless, because they can be prepared and maintained relatively easily, their neurological condition is of some interest

These animals retain their righting reflexes and can regulate their body temperature They are also capable of carrying out coordinated reflex acts and show, often to an exaggerated degree, reactions which are associated in the intact animal with the emotional states of fright and of anger Such reactions, termed pseudoaffective, are also exhibited by an animal whose cortex alone has been ablated, the condition known as "sham rage " In contrast to the decerebrate animal, the distribution of muscular tone in the thalamic cat or dog shows little departure from the normal, though pronounced extensor rigidity becomes evident when the animal is held suspended in midair

Primates, on the contrary, are usually unable to walk after bilateral decortication, and there are profound alterations in the muscular tone of the limbs The animal generally assumes a characteristic attitude When lying upon its side the limbs of the under part of the body are extended and the uppermost limbs flexed (Fig 9 80) The uppermost limbs show a pronounced grasp reflex Turning the animal on to its opposite side reverses the picture The limbs which were underneath and extended are now flexed and show the grasp reflex, while the limbs previously uppermost, being now underneath, are extended Righting reflexes (except those dependent upon vision) and the tonic labyrinthine and neck reflexes of Magnus and de Kleijn are shown in a striking fashion Fig 9 81

If the extirpation of the cerebral cortex of a monkey is carried out piecemeal over a considerable period of time, and if the animal is

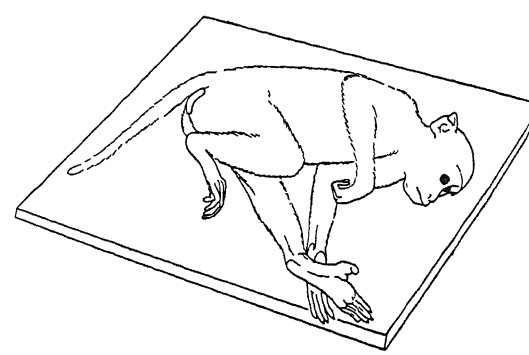


Fig 9 80 Posture of decorticate monkey when it is lying on its left side The two lower limbs are extended the upper ones are flexed

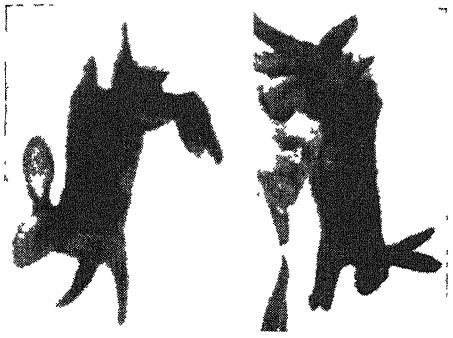


Fig 9 81 Posture of decorticate rabbit suspended from a hand grasping the pelvis *A* The labyrinths are intact and there is righting of the head and neck with postural response in forelimbs *B* After labyrinthectomy the head and forelimbs hang limp (Magnus 1926)

given meticulous care as to nutrition and manipulation of its extremities, it may be able to stand briefly and awkwardly and may even take a few steps, but in general such an animal is quite disabled The disability is much more profound than that seen in a decorticate or thalamic cat

Reflex Maintenance of Posture

By observing cats and rabbits after operations upon the central nervous system, Rudolph Magnus (1926) and A de Kleijn (1923) of Utrecht were able to describe the responses through which posture is maintained and righting achieved (Their papers are in English) They classified these responses according to what they do for the animal, whether righting, change in posture, movement of the eyes, etc , and also in terms of the sensory organs from which the responses originate The most important of these sense organs include *a*) the visual system, *b*) the semicircular canals, utricle, and saccule of the labyrinth, *c*) stretch receptors in muscles and ligaments of neck muscles and joints, *d*) pressure receptors in toe pads and in other surfaces of the body, and *e*) stretch receptors of muscles of the limbs The function of these responses is to govern the orientation of the head in space, the position of the head and eyes in relation to the trunk, and the position of the limbs and eyes in relation to position of the head They aid in maintaining the orientation of the body in relation to gravity, as well as restoring the body to an upright position when its posture has been disturbed Some of these phenomena are true reflexes, whereas for others the word, response, is more appropriate The nature of the response usually is evident from the name that has been given to it

A) LOCAL AND SEGMENTAL STATIC REACTIONS

1) Positive Supporting Reaction simultaneous reflex contraction of both extensor and flexor muscles that converts the limbs into rigid pillars It is initiated by stimuli from pressure receptors, e g , in sole of foot, and proprioceptors in muscles that would be stretched if the leg collapsed

2) Negative Supporting Reaction relaxation that permits the limb to be moved to a new position It is initiated by passive flexion of the limb

3) Segmental Static Reactions flexion and crossed extension reflexes, are described in chapter 4

B) GENERAL STATIC REACTIONS

1) Statotonic or Attitudinal Reflexes

a) *Tonic Labyrinthine and Tonic Neck Reflexes Acting upon the Limbs* influence tone of skeletal muscles and maintain the different parts of the body in an attitude appropriate to a given position of the head Receptors are in the labyrinth (utricle and saccule) and in muscles, tendons and ligaments of the neck In an animal on all fours the labyrinthine effect is to increase the extensor tone in the muscles of all four limbs when the head is strongly extended, and to decrease the tone when the head is flexed The tonic neck reflexes, by contrast, evoke flexion of the fore limbs and extension of the hind limbs when the neck is flexed towards the sternum, extension (dorsiflexion) of the neck produces the converse picture Rotation of the head causes increased tone in the fore- and hindlimbs on the side of the body toward which the jaw is rotated, and reduces the extensor tone in the opposite limbs. In man this is the classic pose of a boxer.

b) *Tonic Labyrinthine and Tonic Cervical Reflexes Acting upon the Eyes* tonic effects upon eye muscles, anal-

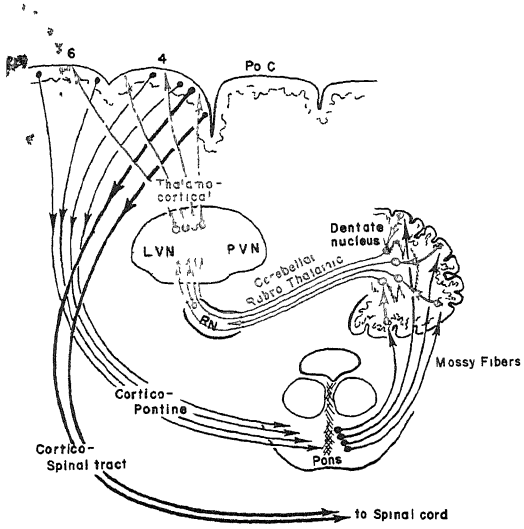


Fig 9 82 Pathways by which the cerebral cortex pons cerebellum red nucleus and thalamus are interconnected Interruption of this circuit between cerebellum and thalamus may cause an intention tremor

produces smooth movement becomes oscillatory when one component, the cerebellum, is disconnected But the subsequent removal of another component, the premotor cortex, decreases the oscillations Perhaps this second operation decreases the power transmitted through the system, or at least through that part that previously evoked the oscillation This topic is considered later in this chapter in connection with rhythmic activity in the central nervous system

The deficiencies in motor movement after removal of the cerebellum include a prolongation of the latency of response of both contraction and relaxation of muscles, together with both increase and decrease in strength of individual muscle contractions All of these effects represent dysfunction of the corticospinal systems (including their γ motor portions), and have been utilized for a recent hypothesis as to how the cerebellum functions (Blomfield and Marr, 1970) Eccles et al (1971) have suggested that the key to cerebellar function is found in the distribution of connections to and from the Purkinje cells

There are many experiments that suggest that the influence of the cerebellum on the corticospinal system is facilitatory, whereas others imply an inhibitory effect. The central nuclei of the cerebellum apparently belong functionally to the bulbar facilitatory system, as noted earlier, so that stimulation strong enough to activate them will increase muscle tone and strength, especially in the extensor muscles of a cat The cortex by way of the Purkinje cell projection, by contrast, inhibits central nuclei and Deiter's nucleus of the brainstem (Fig 9 83) Cortical stimulation alone, therefore, should decrease muscle tone, especially in the extensors The discovery of these two opposed types of actions should lead eventually to a better understanding of data from stimulation experiments At the present time such data are contradictory, perhaps because stimuli applied to the cortex would not produce any apparent effect because the resulting inhibition could be observed only against a background of antecedent activity

FLOCCULONODULAR LOBE AND OTHER PARTS

Through its vestibular connections the flocculonodular lobe is concerned with equilibratory function, severe disturbances of balance resulting from its ablation Removal of this lobe in monkeys, or of the nodule and part of the uvula, is followed by pronounced disturbances of balance In man a lesion of this part of the vermis is also associated with disturbances of equilibrium Removal of the *nodule* alone in monkeys is followed by oscillation of the head and neck (and, in some instances of the whole trunk) falling backwards and an unsteady gait with the extremities spread widely apart to provide a broader and thus more certain base Ablation of the nodule in dogs confers immunity to motion sickness Removal

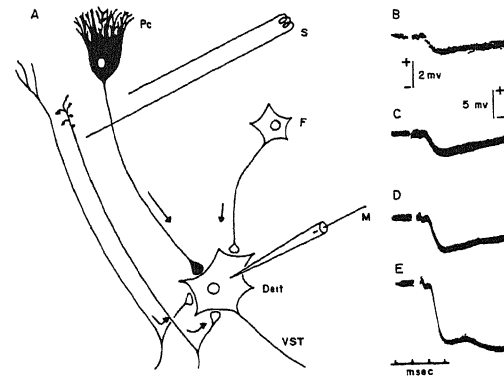


Fig 9 83 Demonstration of inhibitory action of a Purkinje cell upon a cell in a vestibular nucleus (Deiters) *F* is an excitatory cell from fastigial nucleus, *VST* is vestibulospinal tract *S* is stimulating electrode *M* is the intracellular, recording electrode The records *B–E* on the right show the development and course of the inhibitory postsynaptic potentials (*IPSP*) (Eccles 1967)

of other regions of the cerebellum or of the cerebral cortex does not have this type of protective action

The functions of the *uvula* are largely obscure, its destruction alone is followed by some transient disturbance of equilibrium Very little is known of the function of the *pyramid*, although it appears to be concerned in some way with vision or eye movement Its excision causes no defect of balance Ablation of the pyramid in the monkey results in an inability to halt its forward progression in time to prevent it from crashing into some obstacle Faradic stimulation of the pyramid elicits an upward movement of the eyes, and Fulton argued from this fact and the preceding observation in favor of the possibility of the pyramid's being concerned in some way with the integration of proprioceptive impulses essential in judging distance

Basal Ganglia

The many connections and interrelations the basal ganglia have with other parts of the somatic motor systems (chap 2) suggest that they must participate in a considerable variety of motor functions (Fig 9 84) What is known of their properties can be summarized in two statements, as follows *1*) they stabilize or "damp" movements that otherwise would be rhythmic or oscillatory, *2*) especially in lower animals but presumably also in man they have the capacity to initiate movement

It has been difficult to attribute specific functions to individual ganglia, in part because lesions created experimentally usually do not produce obvious deficits Perhaps this is because the requisite imagination and insight have not yet been brought to bear on the problem of testing the performance of the animals If one is to disclose an input-output relationship, the input must be appropriate for the mechanism being analyzed

DYSFUNCTION OF BASAL GANGLIA

Information about functions of the basal ganglia has come principally from observation of patients with lesions in these structures Three types of lesions are typical, as follows

1) Lesions found in the putamen of the lentiform nucleus as a result of copper intoxication in Wilson's disease—hepatolenticular degeneration—cause tremor and rigidity

2) Lesions found in the substantia nigra following manganese intoxication cause symptoms of Parkinson's disease (paralysis agitans), as described under 3)

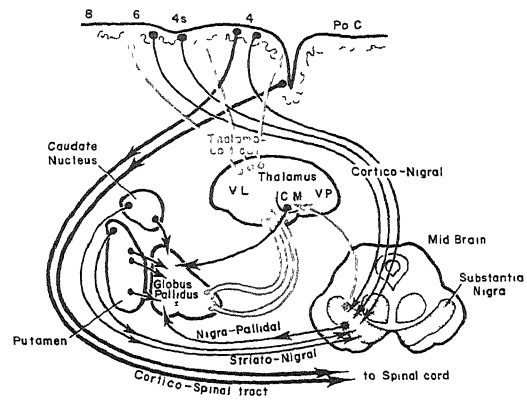


Fig 9 84 Portrayal of some of the better understood interconnections of the basal ganglia brainstem thalamus and cerebral cortex

3) Lesions that remain following encephalitis, and that are believed to occur also in patients with cerebral vascular disease, are responsible for paralysis agitans They involve the globus pallidus and substantia nigra, with depigmentation of the latter because of a failure of melanin accumulation

In addition to these three, other lesions may occur as a result of vascular disease, infections, or the growth of tumors

The symptoms of paralysis agitans include both slowing and weakening of somatic motor movements, loss of "associated" movements, for example in walking (see later), tremor of postural mechanisms, rigidity to passive manipulation of limbs, and an apparent dulling of emotional responses The tremor is known as a *tremor at rest,* it involves typically the muscles of the neck and upper limbs By contrast, the tremor that accompanies voluntary movement in patients with lesions of the cerebellum or its projection to the lateral ventral nucleus of the thalamus is an *intention tremor* Other symptoms associated with disease of the basal ganglia include two forms of spontaneous and involuntary movements, *chorea* in which the movements are quick and jerky, although apparently coordinated, and *athetosis* where they are slow, sinuous and writhing, and involve particularly the hands and forearms.

CONTROL FUNCTIONS

Although it has been repeatedly said that there is no localization of function, somato-

topic or otherwise, in the basal ganglia, clinical experience leaves no doubt that this statement is untrue As noted above, disease in different parts of the basal ganglia gives rise to vastly different clinical pictures Destruction at any point in the dentatorubrothalamic projection (Fig 9 82) causes an intention tremor Lesions of the substantia nigra and globus pallidus are associated with parkinsonism and a tremor at rest Destruction of the caudate nucleus and putamen is seen in patients with choreoathetosis Obviously there must be some functional differences between these various nuclei Every clinical neurologist is aware of how the tremor of parkinsonism may begin in the head and neck, or in the foot, or in a finger, and then later gradually spread to other parts of the body on the same or opposite side Obviously there is somatotopic localization in the basal ganglia in man, otherwise such localized manifestations of disease would not be possible

Voluntary Muscular Activity. For years it was supposed that voluntary muscular activity originates exclusively in the corticospinal system It is now obvious that that is incorrect, at least in species up to and including carnivores (see p 9-98) It is true, however, that in primates the situation is quite different Does this mean that in primates the basal ganglia and other subcortical centers have lost their ability to produce useful voluntary movements? This seems unlikely, but it does appear that they have lost their ability to produce such movements in the absence of the precentral motor cortex It may be that when the cortex of a primate is removed the basal ganglia are left in a hypoactive condition analogous to spinal shock, and thus are incapable of initiating movement If, on the contrary, these ganglia become hyperexcitable (as in dopa therapy, see below), then movement is generated on an involuntary basis

Reflex Muscular Activity. Rigidity is one of the characteristics of disease of the basal ganglia, notably Parkinson's *paralysis agitans* Clinicians frequently attempt to make a sharp distinction between the rigidity of diseases of the basal ganglia and the spasticity associated with decerebration and capsular hemiplegia However, the differences appear to be rather superficial As Fulton (see 1949) and his coworkers were the first to emphasize, both result from the loss of supraspinal inhibition which is not concerned with the corticospinal fibers Furthermore, the rigidity of paralysis agitans, like the spasticity of decerebration, is greatly diminished by interrupting afferent impulses from the involved muscle to the spinal cord This demonstrates that both rigidity and spasticity result primarily from activity of the local, spinal reflex arc which is released to hyperactivity by removal of descending, inhibitory influences Thus, it appears likely that the basal ganglia as well as the precentral motor cortex, area 4s and the bulbar reticular formation, exert an inhibitory influence over the spinal reflexes which control the muscular activity responsible for posture and reflex standing Clinical observations that lesions in the globus pallidus produced surgically in the treatment of various diseases result in the reduction of rigidity indicate the possibility that the basal ganglia exert an excitatory as well as an inhibitory influence upon the spinal reflex mechanism responsible for rigidity However, the physiological explanation of these clinical observations is as yet unknown This serves further to emphasize our ignorance in this field

Automatic Associated Movements. One thing that characterizes parkinsonism perhaps even more than tremor or rigidity is the loss of automatic associated movements Commonly this term is regarded as indicating only the swinging of the arms in walking—a type of movement which is abolished by various destructive lesions in the precentral motor cortex, in the cerebellum, as well as in the substantia nigra and globus pallidus in parkinsonism However, in parkinsonism far more is lost than automatic swinging of the arms There is a general poverty of movement The loss of expressional movements of the face results in the familiar masklike face There is also a generalized loss of movements throughout the body so that the victim is described as being "statuesque " It thus appears that the basal ganglia are responsible for a wide variety of automatic and associated movements

CHEMICAL TRANSMISSION

Further understanding of the performance of two of the basal ganglia, the substantia nigra and globus pallidus, has come from recent success in relieving the symptoms of patients with parkinsonism (Figs 9 85 and 9 86) by administration of the amino acid, L-dihydroxy-



Fig 9 85 Chilean ex-miner with chronic manganese poisoning is trying to lift his arm and extend his fingers From Mena and others cited by Cotzias 1969



Fig 9 86 Patient shown in Fig 9 85 during therapy with levodopa Note also change in facial expression Mena et al cited by Cotzias 1969

phenylalanine (dopa) (Cotzias, et al, 1967, 1969) The efficacy of this amino acid in counteracting these symptoms indicates that its derivative, dopamine (dihydroxyphenylethylamine), produced in the nervous system by the enzyme dopa decarboxylase, is an essential transmitter somewhere in this system (Fig 9 87) Two earlier observations had suggested that this type of transmission might be involved First, dopamine was identified in this part of the brain by Carlson and his colleagues (1958), who found also that the concentration was lowered following administration of the drug, reserpine When dopamine was then given after reserpine it did not enter the brain to restore the usual levels, indicating that it cannot pass across the capillary-meningeal boundary that is called the "blood-brain barrier" (Carlson et al, 1957) The administration of dopa, by contrast, did restore the dopamine level in the brain via the decarboxylation reaction mentioned above

Second, Hornykiewicz (1963) found reduced concentrations of dopamine in the substantia nigra of patients with Parkinson's disease This observation is believed to have some as yet unexplained relation to the depigmentation of this structure in these patients Melanin, the pigment, is likewise derived in the body from dopa The metabolic significance of the pigment in these cells, however, is not clear, nor is the reason for the concurrent deficiencies of melanin and dopamine

In the course of treatment of these patients, two other phenomena were noted that are of neurophysiological significance First, inasmuch as most of the dopa given was distributed throughout the rest of the body, (in addition to the fraction that entered the brain,) there was a generalized elevation of dopamine levels as a result of decarboxylation in other tissues, for example, the liver This led to symptoms of dopamine intoxication, and in some cases necessitated a reduction in dose of dopa administered The intoxication was especially marked in early experiments where the racemic or D,L form of the amino acid was the only form available in the quantities needed, when the L form became available the total dose could be cut in half and general symptoms of intoxication were less troublesome Cotzias has suggested (1969) that it may prove feasible to block the decarboxylation reaction in peripheral tissues by giving a blocking drug that, itself, cannot enter the brain.

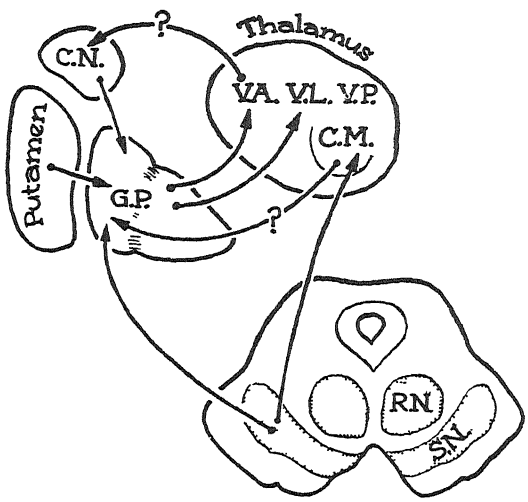


Fig 9 87 Projections from substantia nigra as significant components of interrelationships of basal ganglia Compare with fig 9 84

Second, another type of undesired side effect of the dopa treatment was involuntary movements that included myoclonus, athetosis, and chorea The therapy that relieved the tremor and rigidity, therefore, and that created a sense of "awakening" in the patients, also led to movements that presumably were generated in the basal ganglia

RHYTHMIC MECHANISMS

A significant feature of the involuntary movements associated with disease of the basal ganglia or cerebellum is repetition Steadiness of *a*) normal *posture* is replaced by a tremor at rest, smoothness of *b*) *movement* is obscured by intention tremor, the quietness of *c*) normal *resting* is disturbed by chorea or athetosis Each of these seems to be a condition where the interruption of some neural mechanism "releases" an underlying rhythmic mechanism from a control that is usually superimposed over it.

There is reason to believe that such rhythms are, in general, a fundamental pattern of response of a control system (Wilson, 1967) If a given system does not show such a rhythm or oscillation, it means that some component has been added so as to achieve a smoothing of the response One should recall that the usual condition of a simple system with a negative feedback is one of oscillation (see Grodins, 1963, p 96 ff) As the output increases and the feedback becomes stronger, the feedback is returned as a negative input that lessens the output, with the result that the feedback likewise becomes less intense This permits the cycle to start again and an oscillation follows If the oscillation is not in turn controlled or "damped," and if time relations are what is called "tuned," the oscillation may become progressively greater so as to destroy the system One way of preventing such oscillation is to detune it in some way, or to introduce secondary feedback loops having time constants different from the first one Another way is to add feedbacks having detectors that signal, not where the system is, but where it is going, or how fast it is changing in a given direction Neural circuits having these properties seem to be responsible for smoothness and stability of movement in the somatic motor system, and these concepts seem to offer a useful hypothesis for thinking about involuntary movements If the concepts are correct, one can see that although it might be possible to remove a focus from which an unwanted movement such as chorea or athetosis originates, contrariwise a tremor of resting posture (in parkinsonism) or of voluntary movement (intention tremor) can be relieved only by altering the phase relationships of the system, or actually interrupting the feedback Or, as noted earlier, the magnitude of the oscillation, but not the oscillation per se, may be reduced by decreasing the power applied to the system

Rhythms may originate at any level of the nervous system For example, a single detector creates a rhythmic pattern of nerve impulses, as does a motor neuron, the most fundamental of all repetitive activity Its frequency depends upon intensity of stimulation, sensitivity of the detector or the postsynaptic membrane, and duration of the refractory period This phenomenon has been designated as "a rhythmic response to a steady state" (D H Barron, personal communication) It is responsible for signal generation, as described in chapter 1. If it occurs in the central nervous system it may account for what appears to be "spontaneous" activity including some voluntary movements

When circuits or loops of neurons function together, the possibility of another type of rhythm is introduced, a second order type in which bursts of nerve impulses acquire a pat-

tern that is a frequency modulation of the original train It is this second order of rhythm that underlies the involuntary movements just considered, as well as the electroencephalogram (see next chapter) and rhythmic activity in other levels of the neuroaxis, as follows

1) On a purely reflex basis there is the phenomenon of clonus (p 9-97), the repetitive contraction of a muscle after its tendon is stretched (Fig 9 88) It appears especially in extensor muscles of decerebrate preparations when the myotatic reflexes are hyperactive (Creed et al , 1932)

2) More highly organized is the rhythmic stepping that Sherrington (1906) and many other authors have studied in spinal animals Sherrington found that it did not depend upon innervation from muscles involved, nor upon any other one type of sensory stimulation, and so concluded that it originates in the spinal cord It is presumed to be brought about by nerve nets that are self-exciting in a fashion that alternately stimulates and inhibits both flexors and extensors in a reciprocal fashion in both hind limbs

3) From the brainstem there arise the rhythmic trains of impulses that maintain pulmonary ventilation, as well as another form of rhythm that sometimes appears in blood pressure records as Traube-Hering waves (sect 3, chap 9)

4) Rhythms of feeding and drinking, as well as those associated with temperature regulation and with certain phenomena of reproduction such as the estral and menstrual cycles come from the hypothalamus and pituitary gland

5) The levels of basal ganglia and cerebellum are represented both by the presence and absence of movements described in this chapter

6) Finally, many of the voluntary movements that are believed to begin in the motor cortex are rhythmic in nature, including writing, typing, dancing, and a multitude of others

In certain of these examples it is possible to identify the negative feedback pathway that maintains the oscillation For clonus it is muscle spindles stretched when the muscle

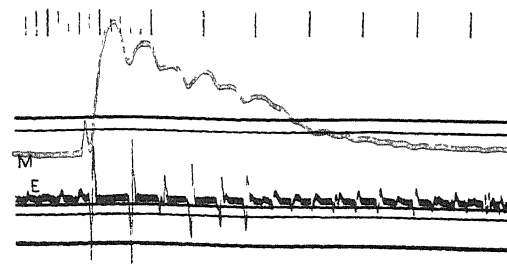


Fig 9 88 Clonus in lightly stretched quadriceps muscle of decerebrate preparation Clonus was induced by tapping the tendon of the muscle Rhythmicity is shown both in mechanical record (*upper*) and electromyogram (*lower*) (Creed et al 1932)

relaxes and the limb moves passively In a cat with a low decerebration the rhythm of pulmonary ventilation may be created by stretch receptors in the lungs that serve to limit each inspiratory excursion If the decerebration is still lower, however, and the vagal nerves are both cut, another variety of rhythm may be seen that is analogous to the stepping of Sherrington's spinal dog after the afferent nerves had been cut This type of rhythm comes from circuits, from negative feedbacks, that are within the nervous system and do not require a peripheral detector They may or may not include a pacemaker, a master neuron that always fires first But they must include connections such that activity eventually is fed back as inhibition that suppresses the original output

The tremor at rest of a patient with lesions of the basal ganglia, therefore, indicates that postural mechanisms are normally rhythmic when cortex, cerebellum, brainstem, and spinal cord are intact and when voluntary motion is not in progress It is the function of substantia nigra and globus pallidus to mask this rhythmicity so as to maintain a stability of posture Similarly, the movements that originate in motor cortex on a "voluntary" basis show an obvious and disturbing rhythm unless they are smoothed out by the circuit that includes the cerebellar cortex What is known about the organization of this part of the brain is consistent with the interpretation that a phase shift is introduced by the more circuitous pathway from cortex to pons to cerebellum and back to thalamus and cortex

Organization and Activity of the Cerebral Cortex

The human cerebral cortex (or *pallium*) has a total area of about 2,200 cm ^{2}, not more than a third of this lies upon the free surface or crown of the convolutions The remaining two-thirds of the gray mantle of the cerebrum occupies the walls and depth of the sulci The total number of nerve cells in the human cerebral cortex has been estimated at around 14×10^9 The number of fibers received from and projected to lower levels of the nervous system is in the neighborhood of 200 million Added to these are fibers, many times more, which connect cells within the cortex (association fibers) Because many of these fibers are not myelinated, their courses are not easily traced from one region to another Of those that are myelinated, a prominent number join the cortices of the two sides via the corpus callosum ("hardened body") that is a major portion of the white matter in the center of each hemisphere

FINE STRUCTURE OF THE CORTEX

When the grey mantle of the cortex is cut transversely, the fine structure is seen to be organized in layers or laminae (Fig 9 89) These layers do not show identical histological appearances throughout the extent of the cortex characteristic differences in the depth of the individual layers and their cellular components are found in the various regions The six layers from the surface inwards are shown in Figure 9 89, as they were named by von Economo they include the following

I Molecular (or plexiform) *layer,*
II External granular layer,
III Pyramidal cell layer,
IV Internal granular layer,
V Ganglionic layer (internal pyramidal layer)
VI Fusiform cell layer

It should be emphasized, however, that even the number of layers is not a feature common to the entire cortex In man one-twelfth of the cortical area shows fewer than six layers (typically three), this portion, the *allocortex* or *archipallium*, comprises the cortex of the olfactory lobe and limbic system (chap 8, p 9 156), and includes the pyriform area and the hippocampal, supracallosal, cingulate, and olfactory gyri The six-layered cortex, which in man constitutes the remaining eleven-twelfths, but which in lower mammals is a much smaller fraction of the whole, is called the *isocortex* or *neopallium*

The greater part of the human laminated cortex shows the cytoarchitectural features described above, and is therefore sometimes referred to as the *homotypical* cortex In other more restricted areas the cortex departs from the typical cytological appearance, chiefly in the preponderances of the small granular cells or of the pyramidal elements, and is called *heterotypical* In areas in which the internal granular layer is absent or practically so, and in which pyramidal cells predominate, the cortex is called *agranular*, it tends to be thick because of the presence of the larger types of neurons When the small, granule-like cells form a prominent internal granular layer with few pyramidal cells, it is called the *granular* cortex, this type is particularly thin at the frontal and occipital poles of the hemispheres

LOCALIZATION OF FUNCTION IN THE CEREBRAL CORTEX

Many authors have tried to make some significant correlation between fine structure and function of the several cortical areas These attempts have not been particularly helpful, except in two instances One is in the case of the cortex of the precentral region (areas 4 and 6), which is of the agranular type The granular layers (II and IV) are shallow, being encroached upon by the expansion of the pyramidal layers III and V In the deep part of layer V of area 4 (the area gigantopyramidalis) are situated the characteristic giant cells of Betz

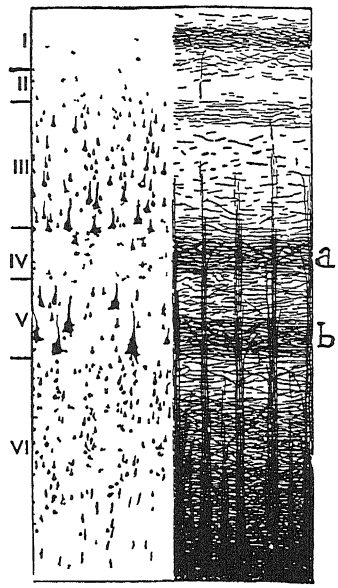


Fig 9 89 Microscopic appearance of cerebral cortex (area 4) as stained for neurons (*left*) and for fibers (*right*) Layers are identified with large cells of Betz in layer V Outer and inner bands of Baillarger are indicated as *a* and *b*

(60 to 80 μm, in their greatest diameters) (see below) These cells contribute fibers to the corticospinal tracts In area 6 rostral to area 4 the granular layers are also inconspicuous even though the large cells of Betz are absent

In the cortex of the walls of, and surrounding the calcarine fissure—the *area striata* or *visual area*—the outer and inner granular layers (II and IV) are expanded at the expense of the pyramidal layers (III and V) This very finely granular type of cortex, also called from its "dusty" appearance under the microscope the *koniocortex*, is characteristic of sensory areas It is present, though to a less fully specialized degree than in the visual cortex, in the postcentral gyrus (somesthetic area) and in Heschl's gyrus (auditory cortex)

Campbell (1905) mapped the cortex into 20 areas which can be differentiated on the basis of their cytoarchitectural characters This number was raised to 47 by Brodmann (1909), and further subdivided by Vogt and Vogt (1919) to make a total of 200 Bailey and von Bonin (1951), however, in a careful reappraisal of the matter concluded that the only significant subdivisions are the following

a) *Allocortex* of the hippocampal region and the pyriform lobe,

b) *Agranular cortex* of the precentral region and the anterior limbic area,

c) *Koniocortex* of the primary sensory areas (the somesthetic areas of the postcentral gyrus, the visual cortex of the calcarine area, and the auditory cortex of Heschl's convolution on the superior surface of the first temporal gyrus),

d) *Eulaminate cortex* of most of the remainder of the mantle of the cerebral hemisphere

Differences, most of them minor, within this broad outline are demonstrable For example, the agranular cortex of the anterior limbic area partakes of some of the characteristics of the allocortex (see chap 8, p 9-156) The koniocortex of the visual area is more complex than that of the somesthetic and auditory areas, and there are minor variations in different parts of the eulaminate cortex which are not of significance here In the following discussion, therefore, the numbers assigned to cortical areas by Brodmann will be used merely for convenience in identifying cortical loci (Fig 9 90) The principal gross anatomical subdivisions of the cerebrum are shown in Figures 9 91 and 9 92

ELECTRICAL ACTIVITY OF THE CORTEX

Electrical activity in the nervous system is of two types, evoked or spontaneous *Evoked* activity is elicited by applying an appropriate stimulus to a sense organ or to the nervous system itself and recording the resultant electrical activity with either a gross electrode or a microelectrode (Landau, 1967) Examples of such activity are shown in Figure 9 93 *Spon-*

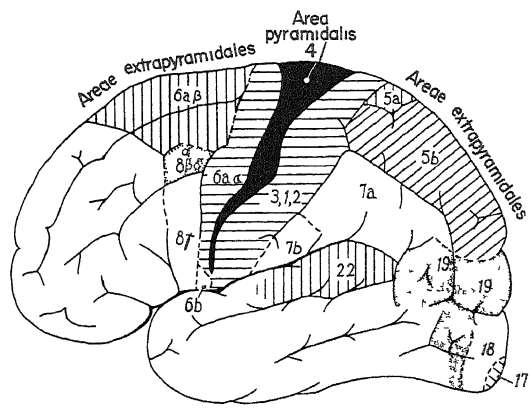


Fig 9 90 Lateral view of the human cerebral hemisphere showing the numbers assigned to the several areas After Brodmann and the Vogts from Ranson and Clark, 1959

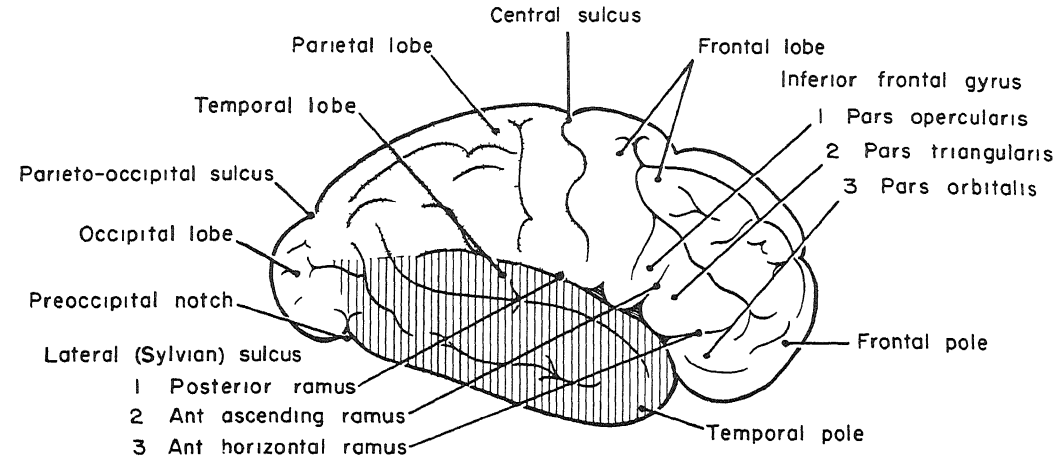


Fig 9 91 Major landmarks of lateral surface of cerebrum Lobes are indicated by shading (temporal) stippling (parietal) or their absence (occipital frontal) From Truex and Carpenter 1969

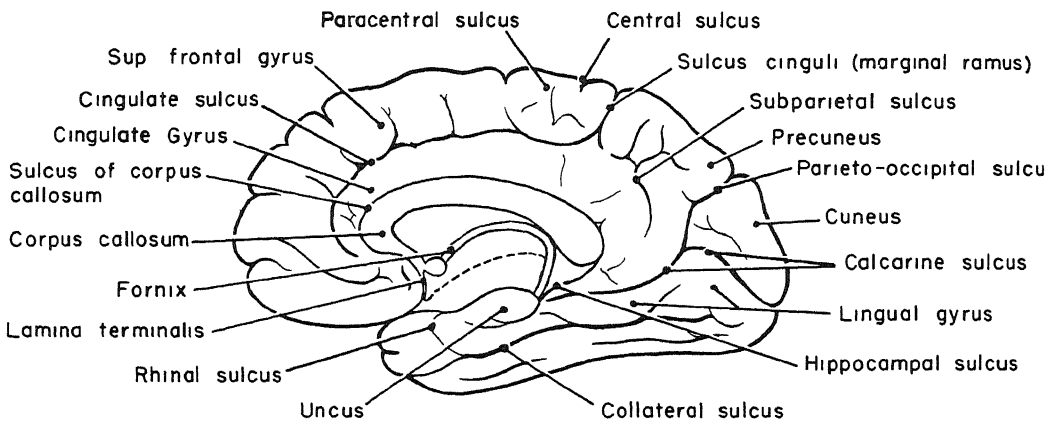


Fig 9 92 Major landmarks of the medial surface of the cerebrum with brainstem cut across and removed below the level of the diencephalon (Truex and Carpenter 1969)

taneous electrical activity is present although the animal is not being stimulated. An example of spontaneous activity is the electroencephalogram, which will be described in more detail later Examples of spontaneous activity are also shown in Figure 9 93 Evoked and spontaneous activity have been subdivided according to whether the activity represents the "spikelike" action potential of the neuron recorded with a microelectrode or a sum of the longer lasting postsynaptic potentials described in chapter 4

The action potential of a neuron in the central nervous system may be recorded either by extracellular or by intracellular techniques For the latter purpose fine electrodes can penetrate a single cell without producing any apparent damage and can record not only the action potential but also the resting potential Both extracellular and intracellular electrodes can record postsynaptic potentials With an intracellular electrode the postsynaptic potentials are derived from the cell penetrated With an extracellular electrode, be it a microelectrode or a larger one, the postsynaptic potentials recorded are considered to be largely composed of the sum of the potential changes of many cells It is this summation that yields the electroencephalogram (Morrell, 1967), and also the slow wave evoked reponses recorded from the nervous system in response to artificial stimulation

The Electroencephalogram (EEG). In 1875 Caton first discovered that an electrode placed on the cortex of experimental animals would record spontaneous fluctuations of electrical

activity Although this discovery was widely reported at the time, the EEG did not receive significant attention for over 50 years In 1929, Berger reported that changes in electrical potential could be recorded from the head of the human subject by means of electrodes applied to the scalp or needle electrodes placed in contact with the periosteum of the skull These brain potentials were later studied by Adrian and Matthews (1934) In normal subjects the EEG may be characterized by three frequency bands, the α, β, and Δ rhythms (A faster rhythm (γ) appears in rare instances) The α rhythm consists of rhythmical oscillations in electrical potential occurring at the rate of 8 to 10/sec The waves have a voltage of about 50 μV on the average The β rhythm has a frequency of 15 to 60/sec and the waves are of lower voltage (5 to 10 μV), the frequency of the Δ waves is from 1 to 5/sec and the voltage is relatively high (20 to 200 μV) No precise cytoarchitectural area can be said to have a characteristic rhythm as was once supposed, but differences are found between rather large areas of the cortex Thus, the α rhythm dominates records from the occipital region or from the posterior parts of the parietal and temporal lobes The sensorimotor region emits waves with a frequency of from 20 to 25/sec, the anterior frontal areas give out waves at the rate of from 8 to 5/sec

The α rhythm occurs in the *inattentive* brain, as in drowsiness or light sleep, narcosis or when the eyes are closed, it is abolished by visual and other types of stimulus or by mental effort (e g , mathematical calculation) The Δ waves can be recorded very rarely from a normal adult while awake, but appear normally during deep sleep or during the waking hours in early childhood Generally speaking, their presence in an adult, except during sleep, indicates some pathological process in the brain—tumor, epilepsy, raised intracranial pressure, mental deficiency, or depression of consciousness by toxic or other factors When present they tend to displace the α rhythm Neither the β nor the Δ waves is affected by opening or closing the eyes

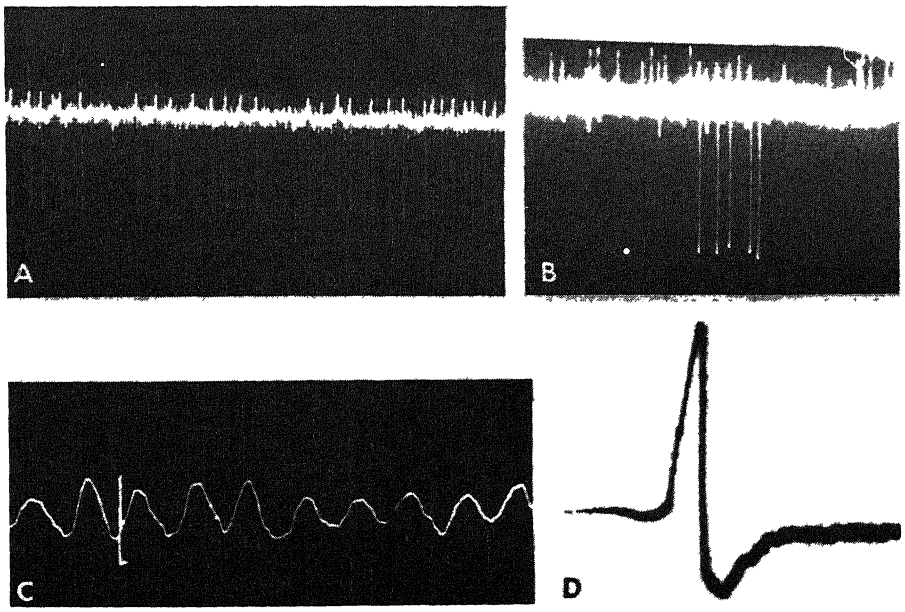


Fig 9 93 Examples of spontaneous and evoked cortical activity *A* spontaneous single unit action potentials recorded from cerebellar cortex with a microelectrode The 'spikes represent the action potentials of a single neuron *B* Evoked single unit action potentials recorded from primary auditory cortex with a microelectrode The auditory stimulus (a click) was presented at the point in time represented by the dot The time delay between stimulus presentation and response represents the latency of the cortical response *C* Spontaneous slow wave activity recorded from the surface of the scalp with a gross electrode (the electroencephalogram) The subject was awake and relaxed with eyes closed The waves have a frequency of about 10/sec—the α rhythm *D*, Evoked slow wave activity recorded from the surface of the primary auditory cortex with a gross electrode The auditory stimulus was applied at the start of the trace

The EEG is affected profoundly by certain general states *Hypoxia* causes at first a moderate slowing of the rhythm, but as the oxygen lack becomes more severe large Δ waves appear, with persistent and severe hypoxia the amplitude of the waves declines and may reach almost the vanishing point *Hypoglycemia* causes an effect somewhat similar to that of hypoxia, when the blood sugar falls below about 40 mg %, Δ waves (1 to 3/sec) are seen, though the α rhythm is not abolished entirely The effects of hypoxia and hypoglycemia enhance one another, their summed effects being much greater than the effect of each alone Hyperglycemia has little effect upon the cortical potentials Increase in carbon dioxide tension increases the rate, but reduces the amplitude of the waves Blowing off of CO_2, as by overbreathing, has the opposite effect

Abnormalities of the electroencephalogram in brain tumor, especially of so-called silent areas of the brain, and in epilepsy, may be of diagnostic value

However, the electroencephalogram may give misleading information as to the location of a brain tumor and a "normal" electroencephalogram in the presence of a brain tumor is not rare Epileptic seizures (see Chap 9) are characterized by pronounced departures from the normal rhythm In *petit mal* attacks, large slow waves appear about a second before the attack is clinically manifest, and displace the previous rhythm Each large wave is followed by a sharp spike deflection (Fig 9 94) In the tonic stage of a *grand mal* seizure, waves of relatively high frequency (10 to 30/sec) and of low voltage appear, but as the attack progresses into the clonic phase these fast waves give place to slower and large waves which continue into the stage of stupor following the seizure Δ waves may be a prominent feature of the electroencephalogram of epileptics between seizures Approximately one-third of epileptic adults have "normal" electroencephalograms

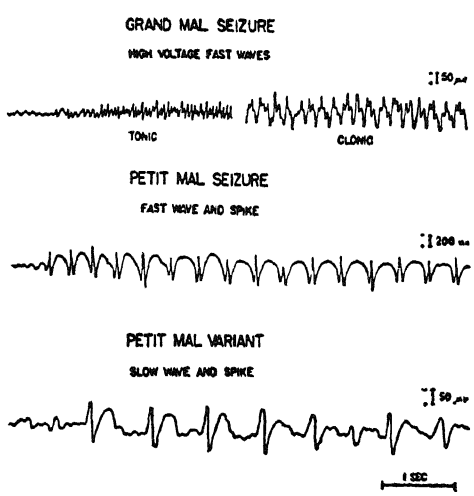


Fig 9 94 Electroencephalograms as recorded during three different types of epileptic seizure

THALAMOCORTICAL PROJECTIONS

Thalamic pathways to the cortex may be placed in two categories the specific and the nonspecific systems (Fig 9 95) Projection fibers of the *specific* system originate in the sensory relay nuclei (e g , the geniculates and the ventral posterior nuclear group) and the association nuclei (e g , pulvinar, lateral nucleus) and, in general, project directly to the cortex Typically these fibers show relatively little lateral spread in the cortex This is particularly true of fibers originating in the relay nuclei Consequently, electrical stimulation of restricted regions in these specific thalamic nuclei results in a relatively discretely localized cortical evoked potential This evoked potential has a latency of 1 to 5 msec If a specific thalamic nucleus is stimulated repetitively at a rate between 6 and 12 pulses/sec , cortical evoked responses to successive pulses may become larger These are called *augmenting* responses and they are limited to the specific sensory cortex

The *nonspecific* thalamic nuclei such as the intralaminar group and nucleus reticularis give rise to a projection system considered to be the rostral portion of the ascending reticular activating system This nonspecific thalamic system is widely distributed over the surface of the cortex, including those cortical areas to which the specific relay nuclei project Many of these nonspecific thalamic nuclei have not been shown to have any direct pathway to the cortex, the conclusion being that these nuclei send fibers to the cortex by an indirect, multisynaptic pathway The exact nature of this indirect pathway remains to be clarified

The specific and nonspecific systems are not independent, but interact with each other at all levels between the brainstem and cortex Many physiological and anatomical studies have shown that collateral fibers from the

somatic, visual, and auditory specific sensory systems enter the diffuse system in the brainstem and thalamus Conversely, fibers pass from the centrum medianum, a nonspecific thalamic nucleus, to terminate in specific thalamic relay nuclei (Nauta and Whitlock, 1954) Electrophysiological studies have shown that stimulation or extirpation of regions of either the specific or nonspecific systems will influence the responses of the other system For example, recruiting responses recorded from primary visual cortex are increased in amplitude after destruction of the lateral geniculate (Hanberry and Jasper, 1953) This increase in the recruiting response recorded at primary sensory cortex has also been seen after destruction of the other major thalamic sensory relay nuclei Moreover, the responsiveness of sensory cortex to specific afferent stimulation is modified by the activity of the nonspecific projections (Fig 9 95) A possible interpretation of these and many other related studies is that the activity of the nonspecific system controls the excitability of the cortex



Fig 9 95 Thalamo-cortical projections The labels show region of origin of diencephalic fibers projecting to cerebral cortex From E Walker as summarized by Penfield and Jasper 1954

Regions with Sensory Functions

How the cerebral cortex, thalamus, and other levels of the central nervous system transmit impulses that lead to sensation is fully considered in section 8 of this textbook Consequently, only a brief account is needed here (Fig 9 96)

SOMESTHETIC AREA OF THE PARIETAL LOBE

The *postcentral gyrus*, i e , the band of cortex lying behind and including the posterior lip and wall of the fissure of Rolando (areas 3, 1, and 2) is sensory in function and is known as the *somesthetic area* This band of sensory cortex turns over the upper border of the hemisphere, and extends down the mesial surface as far as the cingulate gyrus As in the case of the primary motor area, (see later), the sensory areas of the body are, in a general way, represented mainly in inverted order, from the lowest part of the mesial surface to a corresponding part of the lateral surface (see Fig 9 97) The order of this cortical representation conforms closely to that of the dermatomes, i e , to the spinal innervation For example, the cortical sequence for the upper limb (monkey) is postaxial arm, occiput, ear, side of head, neck, shoulder (C2–C6), preaxial arm, forearm, wrist, hand, digits (C7, C8, T1) This dermatomal representation, however, is not in continuous numerical sequence from head to tail For example, if the sensory band of cortex (which turns over the upper border of the hemisphere) could be straightened out, it would appear quite clearly that the projection of the dermatomes from the upper thorax to the tail (thoracic 1 to caudal 4) is in an order the inverse of that of the spinal segments, whereas the representations of the head and neck are not inverted, the cervical segments being represented in ascending numerical order from above down, with the ophthalmic division of the fifth nerve following C8 (Fig 9 97) There is no obvious explanation for this arrangement Excisions of portions of the somesthetic cortex are followed by disturbances of cutaneous and kinesthetic sensations in the corresponding region of the opposite side of the body These subside to a greater or lesser degree with the passage of time

Cushing (1909), and later, Penfield and his associates (see 1950, 1966) stimulated different parts of the somesthetic areas in conscious patients, sensations on the contralateral half of the body were experienced, their locations bearing a constant relationship to the point stimulated.

The sensory area of the cortex is not confined to the postcentral gyrus, but extends forward into the precentral gyrus Penfield and his associates found that 25% of stimulations of the latter area in patients caused a sensation, with or without a motor response On the

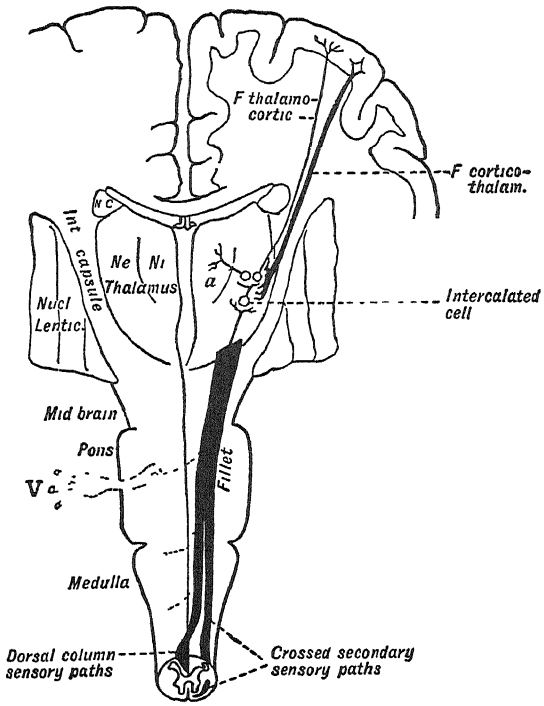


Fig 9 96 Outline of pathways by which sensory impulses ascend from spinal cord through brainstem to thalamus and thence to somesthetic areas of parietal cortex

other hand, 20% of stimulations of the postcentral gyrus gave a motor response, instead of or accompanied by a sensation Moreover, ablations of postcentral areas cause motor defects which are not due simply to loss of kinesthetic sense It appears that the precentral and postcentral areas are knit together by connecting neurons, and are so interrelated functionally as to be called appropriately, taken as a whole, the *sensorimotor area*

The extent of the sensory representation in the postcentral gyrus is a function of the number of sensory fibers reaching the cortex from a given region It is greatest for those parts of the body which are of most importance in acquiring information concerning surrounding objects Thus, in man and in the monkey the cortical area for the hand and arm is larger than that for the trunk or leg A large facial representation in man is in keeping with the generous sensory innervation provided to the face by the three divisions of the trigeminal nerve

Somatic sensation as well as the visual and auditory senses have a double representation in the cortex of each hemisphere This dual somatic representation was first observed by Adrian (1940), in the case of the claws of the cat The somesthetic area described above and long recognized, is now called the "primary" sensory area, or, as Woolsey (1958) has suggested, *somatic area I* The more recently discovered sensory area is called *somatic area II* Corresponding terms for the dual visual and auditory areas are employed, namely auditory and visual areas I and II Somatic area II lies in the upper wall of the Sylvian fissure, that is, below the face area of somatic area I, which is much larger in extent The order of representation in sensory area II is the inverse of that in sensory area I, the face being above the arm in the upper part of the wall of the fissure with the foot at the bottom The significance of the dual cortical representation is unknown

THE TEMPORAL LOBE

The primary cortical center for hearing is situated in the transverse gyrus of Heschl lying in the floor of the lateral cerebral (Sylvian) fissure, and an adjoining small area of the superior temporal gyrus Fibers from the medial geniculate body reach this *auditosensory* area (area 41) via the posterior limb of the internal

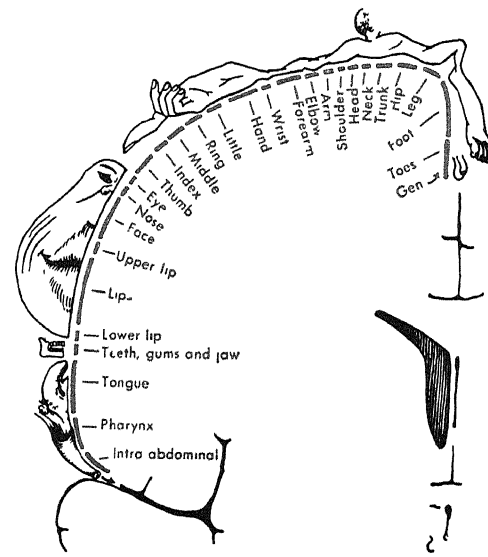


Fig 9 97 Sensory homunculus Diagram illustrates the region of the body from which sensory fibers enter the cortex at each location (Penfield and Rasmussen, 1950)

capsule, they constitute the *auditory radiation* In the auditosensory area the fundamental auditory sensations—intensity, and pitch—are appreciated The area is bilaterally represented Equilibratory sense is represented in the posterior part of the first temporal convolution Stimulation of this region in conscious patients causes dizziness or nausea, a sense of swaying, falling, or of rotation

THE OCCIPITAL LOBE

The gray matter forming the walls of, and surrounding the calcarine fissure (on the medial aspect of the occipital lobe) constitutes the primary cortical center for vision—the *visuosensory area* From the broad stripe of Gennari which can be seen with the naked eye this area is commonly known as the *area striata* or, following the numerical terminology, as area 17 It is considered in more detail in section 8 Its histological features have already been touched upon

The *second visual area* wherein the visual sensations are interpreted and integrated into more complex perceptions is contiguous to the area striata and lies on the lateral aspect of the occipital lobe (area 18)

Stimulation of the anterior part of the lateral surface of the occipital lobe causes conjugate deviation of the eyes to the opposite side (occipital motor eye field, area 19) In man, visual hallucinations, such as, flashes of light of different colors, or definite images have been evoked by the electrical stimulation of area 18 or 19

SENSORY FUNCTIONS OF ASSOCIATION CORTEX

Sensory stimuli produce evoked responses at a number of cortical regions other than primary sensory cortex These regions have been described in the cat as the anterior part of the sigmoid gyrus and the suprasylvian gyrus (Albe-Fessard and Rougeul, 1958). In the monkey, the corresponding areas have been described as superior temporal cortex, superior frontal cortex and parietal association cortex (area 5) (Albe-Fessard et al , 1959) These responses are particularly evident in the unanaesthetized animal or in animals anaesthetized with chloralose They are more labile than primary sensory responses, showing wide variations in magnitude of response to repetitions of the same stimulus When stimuli are repeated, responses in association cortex will fail to appear at every stimulus if the repetition rate is higher than approximately 5/sec On occasion several seconds or minutes may be required for full recovery of response amplitude

One of the outstanding characteristics of association cortex is the diversity of stimuli that converge on these areas For example, a single cortical electrode may record responses from stimulation of many different regions of the surface of the body Localization seems to be poor in contrast to the relatively high degree of specificity seen in primary sensory cortex Stimulation of a variety of sensory modalities will also produce responses at one region Auditory, somatic, and visual stimuli may produce evoked responses from a single electrode location In addition to this overlapping of fields in association cortex, there is also interaction of these sensory modalities (Berman, 1961) Studies of the responses of single neurons with microelectrodes (Amassian, 1954) have shown that a single unit in association cortex will respond to stimulation of any of the limbs or to stimulation of other modalities

Electrophysiological study of the pathways by which sensory impulses reach association cortex have shown that responses remain after ablation of the relevant primary cortical sensory areas (Albe-Fessard and Rougeul, 1958) and after extensive destruction of the mesencephalic reticular formation (Buser et al , 1959) Coagulation of the centrum medianum of the thalamus, however, will abolish the sensory evoked responses in the association cortex ipsilateral to the lesion (Albe-Fessard and Rougeul, 1958) Electrical stimulation of the centrum medianum will also elicit short latency responses in those regions defined above as association cortex On the basis of such evidence, Albe-Fessard concluded that the sensory input to these cortical areas arrives largely via the centrum medianum

Experiments in which chronically implanted electrodes were used to record evoked responses in association cortex in unanaesthetized, behaving cats have shown a number of relationships between amplitude of evoked response and various behavioral activity

(Thompson and Shaw, 1965) Response amplitudes recorded at suprasylvian gyrus as a result of visual, auditory and somatic stimuli (light flash, click, and forepaw shock) are decreased when the animal is active, relative to amplitudes in the resting state Novel stimuli in any of these modalities will reduce the magnitude of response evoked by stimulation in any of these modalities Response amplitude is also decreased when the animal orients toward a stimulus These results indicate that evoked responses in association areas are modified by activity in other parts of the central nervous system

Cortical Motor Mechanisms

PRECENTRAL CORTEX OF THE FRONTAL LOBE

Area 4 is a tapering strip of agranular cortex lying with its wider end at the upper border of the hemisphere, in front of the central fissure (Rolandic), and occupying the posterior part of the precentral convolution A good part of this area is buried, for it covers the anterior wall of the central fissure It also turns over the upper border of the hemisphere and extends down the mesial aspect as far as the cingulate sulcus Area 4 is a center for control of movement It gives origin to an important part of the corticospinal tracts, and also projects to the pons (frontopontine tract), corpus striatum, red nucleus, thalamus, and the subthalamus Area 4 and area 6 just ahead of it are connected by intracortical fibers The muscles of the various parts of the body are represented in area 4 on the mesial and the lateral surfaces, in an order similar to sensory representation in the postcentral gyrus It is, in general, the inverse of that in the body itself, namely, toes, ankle, knee, hip, trunk, shoulder, arm, elbow, wrist, hand, fingers, brow, eyes, face, larynx, jaw, and tongue (Fig 9 98) The parts of the face are not inverted By suitable electrical stimulation of different parts of this area, movements of muscles on the opposite side of the body can be evoked

The careful studies of Woolsey (1958) on the monkey suggest that the distal parts of the extremities, i e , the hands and feet, are represented near the central fissure while the more proximal joints, i e , shoulders and hips, and the trunk, are represented somewhat more

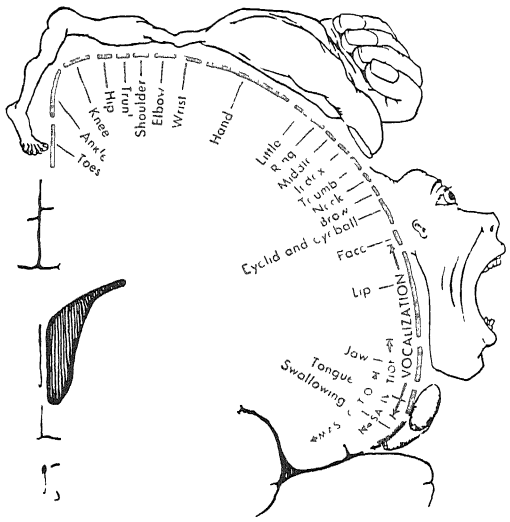


Fig 9 98 Motor homunculus Region of origin of cortical control of parts of the body as determined by electrical stimulation (Penfield and Rasmussen 1950)

anteriorly With suitably controlled stimuli discrete movements limited to a small part of the extremity or even to a single muscle may be elicited Although most of the control of the precentral cortex over the muscular activities of the body is exerted contralaterally, there is also some control over muscles on the same side of the body This ipsilateral innervation is greater in the proximal joints of the extremities than in the distal parts and is greater in the lower than in the upper extremity Furthermore, this ipsilateral innervation by the precentral motor cortex is greater in monkeys than in man

A study by Chang et al (1947), who recorded movements of single muscles, indicates that individual muscles can be stimulated at the cortical level There are, however, certain curiosities about the movements induced by such stimulation Chang and his associates succeeded in recording contractions of single muscles in a reproducible fashion because they strictly controlled both the condition of their anesthetized animals and the circumstances of the stimulation If stimuli are applied more casually to the motor cortex, one finds less constant results For instance, the nature of the motor response obtained from a given electrode placement may change if the frequency or strength of the stimulation is altered Second, the response often depends upon the initial position of the limb, if a limb is passively flexed, stimulation may give extension,

whereas if it is previously extended the same stimulation may yield flexion Third, if the site of stimulation is gradually moved along the gyrus, the response has a tendency to "follow" the movement of the electrode, so that eventually a contraction of arm muscles, for example, may be obtained from what is ordinarily the motor area for leg This shows that there is a difference between motor representation in the grey substance of the spinal cord, on the one hand, and in the cerebral cortex Spinal motor neurons are invariently connected to a given muscle, cortical neurons, on the other hand, seem to connect with various pathways, the nature of the connection depending upon many different factors It is only when all such factors are maintained constant that completely uniform results can be obtained from stimulation of the motor cortex, especially in laboratory primates and in human subjects Granit (1970) has called the population of motor cells that can evoke contraction of a given muscle a "colony"

A *supplementary motor area* (Fig 9 99) on the medial surface of the hemisphere in both man and monkey lies anterior and ventral to the primary motor area described above (Penfield and Welch, 1951, Woolsey, 1958) In it the entire body is again represented The head and upper extremity are represented farther forward and closer to the edge of the hemisphere while the lower extremity is represented more posteriorly and nearer to the corpus callosum The threshold of this supplementary motor area is somewhat higher than that of the primary motor area 4 The function of this supplementary area has not yet been well defined

The *corticospinal fibers* arise from the precentral gyrus (area 4 in the monkey) but not from area 6 (Verhaart and Kennard, 1940) Yet both of these areas do give rise to fibers which descend directly or indirectly to subcortical centers (e g , basal ganglia, thalamus, reticular formation, pontine nuclei, etc) This means that the ablation of areas 4 or 6 is not the same as section of the cerebral peduncle or the medullary pyramid, neither will a cortical lesion give rise to a clinical picture typical of destruction of the motor pathways through the basal ganglia, etc

Neural Deficits Following Cortical Lesions. *Cat or Dog.* Destructive lesions of the various parts of the precentral motor cortex

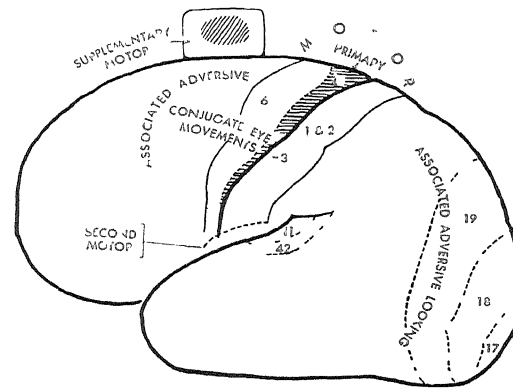


Fig 9 99 Human motor cortex Primary motor area (*4*) is surrounded by areas that are also electrically excitable Supplementary area is on medial surface of hemisphere (Penfield and Rasmussen 1950)

have certain important characteristics If one bilaterally removes the excitable motor cortex (the sigmoid gyrus) from both hemispheres of a dog or cat, or even if all the cerebral cortex is removed, the animal is not completely paralyzed It can sit, stand, walk, or run almost normally, as noted in the preceding chapter The deficits it shows are in the more stereotyped reactions that have been called *placing* and *hopping reactions* These deficits have been analyzed by Bard and Brooks (1934) as follows

Placing Reactions *1*) If a cat is held in midair with legs dependent and chin held up so that it cannot see anything below or in front, contact of the backs of the forepaws with the edge of the table is followed by a quick movement of the limbs which brings the paws, pads down, precisely upon the surface of the table *2*) If the forelimbs of a cat are held down while the chin is brought in contact with the table near its edge, the forepaws when released are instantly raised and placed upon the table beside the chin This movement is usually followed by extension of the limbs and the assumption of a standing position *3*) If the fore or hind limbs of a cat standing or sitting upon a table are pushed over the table's edge, they are immediately lifted and placed in their original positions *4*) If one abducts, without holding, the limb of a standing cat, it is instantly returned to its previous position *5*) If a blindfolded cat is suspended in the air with forelimbs free, and its head brought toward some obstacle, at the instant that the vibrissae come into contact with the object the forepaws

are raised and accurately planted upon the surface of the object The first three of these reactions are due to stimulation of receptors upon the body surface (exteroceptors) and probably also of proprioceptors in muscles and tendons The fourth is a purely proprioceptive reflex, the last is initiated from tactile receptors

Hopping Reactions They involve limb movements that serve to maintain the standing position against any force acting upon it in a horizontal plane When, for example, a cat held so that its body is supported upon one fore or hind limb is then pushed in one or other direction, the supporting limb hops quickly in the direction of the displacement, the foot being kept directly under the corresponding shoulder or hip The hopping reactions are probably dependent upon myotatic (stretch) reflexes

The placing and hopping reactions are controlled from the sensorimotor areas of the cerebral cortex Bard and Brooks found that removal of these regions from both cerebral hemispheres abolishes the placing reactions, and produces an extreme degree of deficiency in the hopping reactions Decortication (complete removal of the neocortex) produces no greater deficiency The control exerted by the cerebral cortex is entirely contralateral, i e , the component movements of the reactions on one side of the body are governed solely by the opposite side of the brain

Cortical Lesions in Monkeys or Man. If the precentral motor cortex is removed from both hemispheres of a monkey, the neural deficit is more severe than in a cat But over a period of months a considerable recovery is seen to the point that, although movements are stiff, visually controlled purposive movement can take place There is, however, permanent loss of what has been termed by Denny-Brown the "instinctive tactile grasping reaction" (1929) This reaction is described as an exploratory palpation which brings an object stimulating light touch into the palm or sole followed by "very facile" grasping Evidence from patients with lesions resulting from vascular accident, surgery, or head wounds indicates that although the ultimate deficit is clearly greater in man than in monkey, a certain amount of recovery of function is possible This recovery is believed to be due, in part, to a reorganization of remaining intact cortex Thus, in monkeys the subsequent removal of parietal (areas 3, 1, 2, 5, 7) or frontal (areas 9 to 12) cortex at some time after recovery from an earlier ablation of areas 4 and 6 increases the pre-existing motor deficit After recovery from localized lesions in hand area 4, hand movements can be evoked by stimulating an adjacent portion of area 4 which had not previously yielded hand movements (Glees and Cole, 1950) However, the ability of totally decorticate cats to move quite well within a short time after surgery indicates that there may be also a subcortical participation in recovery of function In general, the possibility of recovery of function decreases with age and a higher position on the phylogenetic scale Recovery is enhanced by good postoperative care, and training and time

Whereas in the cat and dog the *unilateral* removal of the motor cortex, or even the removal of all the cortex from one cerebral hemisphere, causes little motor disturbance, in the monkey there is an initial flaccid paralysis of the contralateral limbs Within a few days when the animal is again able to stand and run, the extremities contralateral to the hemidecortication are more awkward than the normal ones In jumping the animal may miss the bars of his cage occasionally with the involved digits, and often rests on the backs of the fingers or toes rather than on the palmar and volar surfaces as the normal animal does So long as his normal extremities are available he seldom uses the involved ones to pick up food, but he can do so in case of necessity This means that, although the contralateral extremities are not normal, they are far from being paralyzed Naturally one asks what is the source of the innervation of these extremities There can be little doubt that it is from the intact ipsilateral cerebral hemisphere, for if it is removed, the animal not only becomes paralyzed in the extremities contralateral to the new operation, but also in those affected by the original decortication as well In keeping with this interpretation is the evidence that ipsilateral movements, particularly in the lower extremity and particularly in the more proximal joints could be produced by electrical stimulation of the precentral motor cortex (Holmes, 1909, Bucy and Fulton, 1933)

In man the situation is similar but not identical If the precentral motor cortex is removed from one hemisphere the contralateral extrem-

ities are incapacitated but not completely paralyzed In some cases a slow and rather weak simultaneous flexion of all the digits of the hand can be produced voluntarily Movements at the shoulder and elbow and at the hip and knee are much fuller and stronger In the upper extremity the flexor movements tend to be better, whereas in the lower extremity extension at both the hip and knee is more extensive and stronger than flexion Movements such as these reflect the different functions of upper and lower extremities in man The arm and hand are primarily prehensile, designed to bring things to the individual or to pull him to them The leg, by contrast, is used for standing and walking, primarily as a supporting pillar The fact that the toes and foot cannot be moved is of relatively little importance to a patient who is able to move the extremity at the knee and hip, and so to walk in a very useful fashion, even though he limps and has to circumduct the involved foot with each step Evidence is not available to explain residual or recovered function after cortical injury in man, nevertheless, it seems likely that the ipsilateral cerebral hemisphere plays a role For instance, in a few human cases in which the upper part of the precentral motor cortex has been destroyed bilaterally by a tumor or by injury, both legs have been paralyzed If the movement in the lower extremities seen in hemiplegics were produced by the subcortical centers, then such a paraplegia would not be expected as the result of purely cortical lesions Furthermore, Bates has been able to produce ipsilateral movements by electrical stimulation of the medial surface of the cerebral hemisphere in man

Characteristics of Areas 4 and 6. As noted above, there are differences in function between the various parts of the precentral motor cortex Lesions of area 4 produce more severe and more specific deficits than do lesions of area 6 If area 4 is removed from the precentral motor cortex of monkeys bilaterally, leaving area 6 on one or both sides, the animal is still able to walk, sit, climb, and feed himself but the movements are limited and awkward and reside more about the proximal joints than in the digits (Bucy and Fulton, 1933) Similarly, if the medullary pyramids on the under surface of the medulla oblongata are sectioned in the monkey, the fine, delicate, precise movements of the digits are lost but crude awkward movements of the extremities, particularly those which are more stereotyped and which are related to standing, to progression, etc , are retained (Tower, 1940) Likewise, and somewhat surprisingly, destruction of the corticospinal fibers in the cerebral peduncle in man does not completely paralyze the contralateral extremities, nor is such an operation necessarily associated with the severely incapacitating hemiplegia of a destructive lesion in the motor cortex or in the internal capsule (Walker, 1952, Bucy, 1957) These patients are able to walk and to use their upper extremities for many useful purposes Although final decisive proof is lacking it seems probable that these voluntary movements present in the absence of the corticospinal fibers are produced over the multisynaptic system of the precentral motor cortex and the basal ganglia and other subcortical centers functioning together.

Another technique for studying the function of these regions is to remove all other parts of the precentral cortex and then to stimulate areas 4 or 6 in the absence of the other parts When this is done in a monkey the stimulation of either area 4 or 6 on a single side is capable of producing what appear to be useful movements in any of the four extremities This, again, verifies the ipsilateral as well as the contralateral nature of motor control It further demonstrates that this control may be exerted via the pyramidal i e , corticospinal, pathway from area 4, or via extrapyramidal, multisynaptic pathways from area 6

Only one reflex is clearly altered after destruction of the corticospinal tract from the precentral gyrus If the "foot" area in the uppermost part of area 4 on the medial surface of the hemisphere is destroyed or if the corticospinal fibers to the muscles of the great toe are divided a positive sign of Babinski, dorsiflexion of the great toe, can be elicited by stimulation of the skin of the sole of the foot The normal response to this stimulus is ventriflexion of the toe Destruction of other parts of the precentral region or section of other corticospinal fibers, of course, does not have such an effect

Irritative lesions involving area 4 give rise to localized clonic *convulsive seizures* The convulsive movement usually begins focally in some part of the body depending upon which part of area 4 is irritated There may be a localized twitching of the face, or of one or more

fingers, or of the foot or of the leg This convulsive movement may remain localized or it may spread to neighboring parts of the body or even become generalized This "march" of the convulsive involvement is typical of what is known as a jacksonian epileptic seizure and is dependent upon a spread of the excitation in the motor cortex This subject is discussed in chapter 9, p 9–181

The precentral motor cortex also appears to exert some control over various visceral and vascular activities, such as heart rate, vasomotor activities, gastrointestinal motility, perspiration Stimulation of the motor cortex may affect any of these phenomena under appropriate conditions These changes are said to be similar to visceral and other responses that accompany muscular exercise

Destruction of the precentral motor cortex (areas 4 and 6) results in a spastic hemiplegia, but this spasticity is less severe than that resulting from lesions at lower levels For example, the spasticity following lesions of the internal capsule is in general greater than that seen with cortical lesions In all probability this is due at least in part to some destruction in the basal ganglia, which is common with such lesions, as well as the destruction of the fibers descending from the precentral motor cortex The *pattern* of the increased tone, however, is similar with hemiplegias of both cortical and capsular origin, i e , the increase in tone in the upper extremity is in the flexor muscles With severe lesions at the mesencephalic level it is common to see spasticity even greater than is commonly seen with capsular or cortical lesions but the pattern in both the upper and lower extremities is an extensor one If the lesion is in the spinal cord, one may see either a paralysis with very severe extensor rigidity or a paraplegia in flexion, as has been discussed elsewhere

Grasping Movements and Tonic Innervation. Involvement of the anterior part of the precentral motor cortex, i e , of area 6, is not uncommonly associated with a group of phenomena to which various terms have been applied, e g , "forced grasping and groping," "grasp reflex," "tonic innervation," and so forth A description of these phenomena follows

1) Merely touching the skin between the finger and thumb with a pencil results in slow flexion of the fingers If the stimulating agent is withdrawn gently without disturbing the position of the patient's fingers no tightening of the grasp results but his hand and arm sometimes move through space (grope) in the direction of the moving object, as if drawn by a magnet

2) When an object is placed in the involved hand of the patient the fingers close slowly and gently around it, but any attempt made by the observer to withdraw the object may evoke a grasp reflex in which the object is grasped more firmly Nevertheless, when the patient clenches his empty fist he can relax the fingers again without difficulty

3) Any attempt to bend the arm or leg is met by an active resistance exerted by the antagonistic (stretched) muscles, this differs from the resistance offered by the ordinary spastic limb in that there is no sudden giving way with a "clasp-knife effect "

The gentle grasp and the groping movements just described (see *1*) above) are voluntary and not reflex acts Although they are automatic in nature and are taken to indicate deterioration of the psychomotor functions, the patient can prevent their occurrence if asked to do so They disappear in stupor or in coma The gentle grasping movement follows tactile stimuli alone or visual and tactile stimuli acting together The groping movements can be elicited by visual stimuli alone but not by tactile stimuli alone

These responses have also an involuntary component that depends upon a *tonic innervation* (Walshe and Robertson, 1933), originating as stretch reflexes The strong grasp when an attempt is made to remove an object from the hand is an example of this reflex It is quite distinct from the gentle grasp movement, and results from the passive stretching of the flexor muscles of the fingers as the observer attempts to extricate the object The patient cannot relax the grasp and release the grasped object It occurs in the unconscious patient The grasp reflex is most pronounced when the patient is on his side and the affected arm uppermost

Richter and Hines (1934) produced the "tonic grasp reflex" in adult monkeys by removal of area 6 Excision of the motor areas or the prefrontal areas from both hemispheres did not cause the effect A similar reflex is present in normal infants, it is sufficiently strong to suspend the body from a bar for 2 min A grasping reflex has also been described

for the foot in lesions of area 6, it occurs in the normal infant up to the end of the first year and is elicited by stroking the sole

Other Prefrontal Areas *Area 8*, called the *frontal eye field*, lies in front of area 6 Stimulation of the cortex here causes a conjugate movement of the eyes and turning of the head to the opposite side, opening and closing of the eyelids, and sometimes dilation of the pupils and lacrymation No visual hallucinations occur in man, but an epileptiform seizure may be provoked which spreads to the adversive field (area 6) Ablation of area 8 of one side in monkeys or involvement of this area or its projection fibers in man causes the eyes to be turned up toward the side of the lesion, and paralysis of the conjugate eye movement These effects are temporary The animal in walking will circle toward the side of the lesion No abnormalities are observed in the pupils or in the movements of the lids This area receives fibers from the dorsomedial nucleus of the thalamus, through which it is connected with the hypothalamus It also connects by long association tracts, both afferent and efferent, with the occipital lobe (visual area, 18) and with the tegmentum of the midbrain (probably with the oculomotor nuclei)

Area 44 in the region of the posterior part of the frontal operculum in the *dominant hemisphere* (which is the left in right-handed persons) is the motor area for speech (Broca's area) Stimulation of this area in conscious patients causes their speech to be abruptly arrested Cortical mechanisms of speech are discussed in chapter 8

PREFRONTAL OR ORBITOFRONTAL REGION

The prefrontal region embraces the part of the frontal lobe in front of areas 8 and 44, but it includes the orbital as well as the lateral aspect of the lobe, and for this reason is also known as the orbitofrontal region (Brodmann's areas 9, 10, 11, 12, 13, and 14) These areas were once thought to be inexcitable, and were therefore referred to as "silent" or association areas But upon stimulation by a suitable electric current having pulses of low frequency, autonomic, respiratory, circulatory, renal, and gastrointestinal responses can be elicited After bilateral removal of these areas the blood pressure is reduced, gastric secretion suppressed, and gastrointestinal movements increased In man, suffering from intractable pain is relieved These areas receive important fibers from the dorsomedial nucleus of the thalamus (Fig 9 100), which in turn receives impulses from the hypothalamus These orbitofrontal areas, especially those of the orbital surface, are also closely associated with the "limbic system," as described later

Area 24, anterior part of the cingular gyrus on the mesial aspect of the hemisphere, is a powerful suppressor area Various autonomic responses follow stimulation of this area including pupillary dilation, piloerection, acceleration of the heart, and a fall in blood pressure Arrest of respiration in expiration also results

Unilateral or bilateral removal of the *prefrontal area* does not cause paralysis either in the monkey or in man *Unilateral* removal of a human prefrontal area produces a definite if rather subtle alteration in mental processes Some loss of initiative and mental alertness are the most outstanding results of the operation Memory, judgment and intellect often

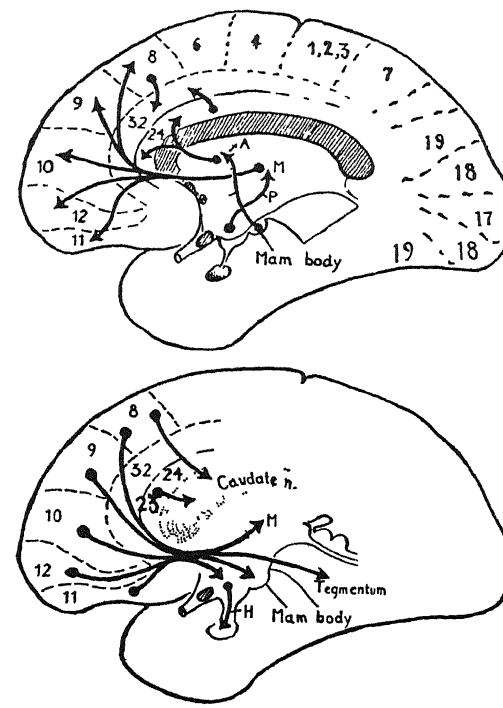


Fig 9 100 Afferent (*upper*) and efferent (*lower*) connections of prefrontal cortex seen from medial view *A* and *M* are anterior and dorsomedial nuclei of thalamus *H* is hypothalamo-hypophysial tract *P* is the periventricular system of fibers

show little or no deterioration Removal of the prefrontal area of the dominant hemisphere (i e , the left in right-handed and the right in left-handed persons) is sometimes thought to produce somewhat greater alterations in character or intellect than does a similar operation upon the nondominant side Even the *bilateral* excision of prefrontal areas is followed by a surprisingly moderate mental defect, so far as superficial observation and examinations are concerned The fact that following a unilateral or even a bilateral prefrontal lobectomy a patient may appear to be normal, however, and functions as a normal individual during casual associations, or passes the usual psychometric tests with flying colors does not mean that the frontal lobes are without important functions so far as mental activity and behavior are concerned If suitable testing is done, it is possible to determine with amazing accuracy the damage to the frontal lobes (Halstead, 1947) Many of these tests are based upon the ability of the patient to classify and categorize a large number of familiar objects and to recall them once they have been taken from him

In an operation for the eradication of a tumor, Dandy excised the frontal lobes on both sides in front of the motor areas (reported by Brickner, 1934) The subject of this extensive extirpation appeared to be of normal intelligence upon a casual acquaintance It is reported that for an hour he toured the hospital with two visiting neurologists who failed to notice in him any mental abnormality A more intimate knowledge of the patient, however, revealed very definite changes of character and mentality His mental age was about 13 years, his intelligence quotient 80 The main features shown by this subject can be taken as generally representative of the effects of extensive prefrontal destruction They constitute what has been called the *frontal lobe syndrome*, as follows

1) *Lack of restraint* leading to boasting, self-aggrandizement, hostility, aggressiveness

2) *Distractibility* and *restlessness*, with difficulty in fixing attention

3) *Hypermotility* which appears to be due to the loss especially of area 13 on the orbital surface

4) *Flight of ideas*, puerile fantasies, emotional instability, facetiousness, punning

5) *Lack of initiative*, and difficulty in planning any course of action

6) *Impairment of memory* for recent events but not for remote events

7) *Impairment of moral and social sense*, loss of love for family

8) *Failure to realize*, or indifference to, the seriousness of his condition, and a sense of well-being (euphoria)

Another manifestation which may follow a prefrontal defect is a disturbance of orientation in time and space The classical example of a severe prefrontal injury which resulted in surprisingly little mental defect, is the case of the American, Phineas Gage (1848), who suffered extensive damage of his frontal lobes by an iron tamping bar driven through his head by an explosion The bar penetrated the left orbit and emerged from the midline of the head just anterior to the coronal suture He was stunned for only an hour and was able with assistance to walk to a surgeon's office He lived for 12 years after the accident and showed in general these mental changes

Chimpanzees which have had both prefrontal areas removed show restlessness and are easily distracted, although they remain alert and evince a keen interest in things around them Fulton and Jacobson reported in 1935 (see Fulton, 1951) that after this operation chimpanzees failed to show temper tantrums and other manifestations of frustration when they made errors in tests requiring the solving of problems or puzzles This observation formed the basis for surgical attempts to relieve certain psychoneuroses in man, first employed by the Portuguese neurologist, Egaz Moniz (1936) The operation, called *frontal lobotomy*, consists in severing the fibers connecting prefrontal areas with subcortical centers (probably the thalamus and hypothalamus) More recently, with the introduction of the various drugs capable of reducing nervous tesnion and rendering an aggresive, agitated patient more tractable, the use of this type of operation has been greatly diminished

Frontal Lobes and Intelligence. The development of the frontal lobes bears in general a direct relationship to the level of an animal in the phylogenetic scale and to its "intelligence " This circumstance has led to the concept that this part of the cerebrum is the seat of the intelligence of animals and the "center" or "organ" of the mind of man Controlled observations, however, indicate that the concept is not valid Following World War II it

was possible to study the effects of traumatic prefrontal injuries in a previously normal population of men with known preinjury intelligence test scores on the Army General Classification Test (AGCT) Administration of the AGCT to these men 10 years after injury showed no decline in their test scores (Weinstein and Teuber, 1957) It seems clear, therefore, that destruction of the prefrontal region does not produce any notable deficit in intelligence

It seems probable that the prefrontal area represents a region of relatively high associative or synthetic capabilities, required for the formulation of abstract ideas and accurate judgment, and for the guidance of conduct in conformity with social customs Here also are mainly located those mental processes relating to "prediction," forecasting, or to any planned action A person deprived of these areas may have little ability as a strategist, thus, a housewife after the loss of these parts of the cortex experiences some difficulty in planning a meal Mental capacity according to this conception is a function of the cerebral cortex as a whole rather than of any particular region (Lashley, 1929)

Lateralization of Function

Because the central nervous system seems to be bilaterally symmetrical in its organization, and also because the cranial and spinal nerves innervate sensory or motor structures on one side of the body or the other, it is of some interest to inquire to what degree the two halves of the system are independent, and where they are necessarily interrelated At the level of primary sensory input and final motor output the two halves are all but autonomous, except for the fact that interconnections occur in spinal cord and brainstem such that the activity of the two halves is naturally integrated For example, ipsilateral flexor reflexes are accompanied by contralateral extensor reflexes (chap 4) Moreover, there are many loci where artificial stimulation yields essentially bilateral responses Thus, Karplus and Kreidl (1910) and also Magoun et al (1938) noted that the autonomic responses elicited by hypothalamic stimulation are bilateral, as are respiratory and other responses evoked from pontile and medullary tegmentum Commissural fibers across the midline of cord and brainstem are plentiful, more than sufficient to account for these results

Similarly, in the cerebrum the commissural fibers from one side to the other are one of its principal anatomical features Most of these fibers run in the corpus callosum, already mentioned as a component of the central white substance of the hemispheres Lesser numbers

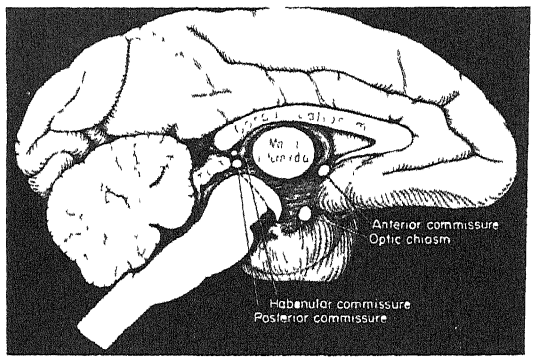


Fig 9 101 Midsagittal aspect of monkey brain showing main structures involved in surgical bisection (Sperry 1962)

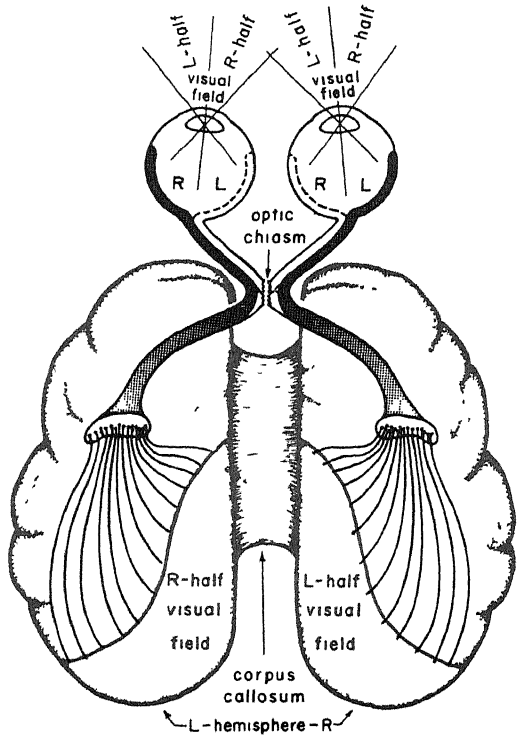


Fig. 9 102 Sketch indicating effects of midsagital division of optic chiasma (Sperry 1961)

are found in smaller bundles, including the anterior and suprachiasmatic commissures rostrally, and the posterior and habenular commissures caudally, with perhaps the massa intermedia that unites the two halves of the thalamus (Fig 9 101)

Some 15 years ago, Sperry and his associates at the California Institute of Technology (see Sperry, 1961) began a study of deficits in animals in which the principal commissures of the cerebrum had been sectioned at surgical operation At the time of initiation of their research the function of the corpus callosum with its multitude of fibers was unknown They were able to study what happens when it is cut because of the imaginative character of the tests they employed, most of them based upon the fact that the visual system is lateralized in its central connections Fibers that leave the retina to form the optic nerve partially decussate in the optic chiasm (sect 8, chap 4) As a result of this partial crossing the right side of the visual field, after registration upon the left half of each retina, is represented in the left lateral geniculate body and striate cortex, whereas the left half of the visual field is carried to the right side of the cortex (The fovea is represented bilaterally) If the fibers that cross in the optic chiasm are cut, and if one eye is then covered, all visual sensation must be achieved in the cortex of the side of the brain on which the eye is functioning (Fig 9 102) This makes it possible to make sure that visual information is reaching only one cortex, depending upon which eye is covered

In their early experiments they discovered (Myers, 1959) that cats tested after section of the optic chiasm and with one eye covered were able to translate visual information to one side of the cortex into control of both sides of the body If the animal learned to perform a certain test with the right eye uncovered, it could do the same test with only the left eye uncovered When the major commissures were then cut at a second operation, however, including the corpus callosum, the animal lost the ability to transfer learning from one side to the other A task learned with one eye open proved to be a new task when presented via the other eye Later experiments on monkeys, baboons, and chimpanzees have in general confirmed the results obtained on cats They have shown, however, that the two halves of the cortex are not completely independent, after section of the corpus callosum and the major commissures either hemisphere is able to control either forelimb for either learned or unlearned behavior, presumably via crossing fibers in brainstem and spinal cord

Similar observations have been made upon a limited number of human patients who have required operations in the region of the corpus callosum Their most notable deficit concerns speech, writing and language, where both the input and the output portions of the systems are lateralized Speech and related activities, such as reading and writing, are represented mainly in the cortex of one hemisphere (chap 8, p 9–158) This has been known for centuries, inasmuch as cortical lesions of one side of the brain may affect speech, whereas similar lesions on the other side leave it undisturbed The usual teaching is that the speech mechanisms are in the cortex of the side that controls the preferred hand—the so-called "dominant" hemisphere For right-handed persons this is the left cortex Penfield and Roberts (1966), however, decided that speech is usually controlled from the left hemisphere no matter which hand the individual uses for skilled actions Whatever the side, it is lateralized

This lateralization of speech makes it possible to test for interconnections between the two halves of the cortex in human subjects After section of the corpus callosum, if the subject with one eye covered is asked to fixate his vision on a point, brief stimuli can then be applied to other positions in his visual field so that information is conveyed to only one of the visual cortices Whether or not he is able to respond to this information via speech, writing, or tests of comprehension is a test of interconnections between the two sides Under these conditions a patient may be unable to name objects that he sees in the left half of his field of vision, the half that is transmitted to the right visual cortex According to Gazzaniga (1970), this type of defect is also exhibited with reference to position sense when a subject is asked to describe the location of his left wrist, fingers, or toes

7 CONTROL SYSTEMS THAT ESTABLISH REGULATIONS

Many sections of this textbook contain descriptions of systems by which the several physiological processes are governed Neural and endocrine control of digestion, for example, is reviewed in the appropriate context (sect 2, chaps 2 and 3) Similarly, the control of circulation and of respiration is outlined where these topics are discussed (sect 3, chap 9, and sect 6, chap 4, respectively) Other examples are found in other sections of the textbook, including almost the whole of the treatise on endocrinology (sect 7) If control systems are as common in physiology as this brief listing implies, perhaps it may be instructive to bring together what is known about them so as to try to discover whether they have any general characteristics (see Yamamoto and Brobeck, 1965, Yamamoto, et al , 1969)

Throughout this textbook the word, *control*, has been used with a meaning practically identical with its general English meaning *Regulation*, on the contrary, is used technically to designate the maintaining of relative constancy of a variable by a system that includes a *detector* for the variable or some derivative of it, and a *negative feedback* of information from the detector to the controller mechanisms (chapt 1, p 9-18)

Typical Regulations

As Figure 9 103 illustrates, the several regulations of the body do have features in common, and can be diagrammed in a similar manner On these diagrams each noun or phrase is a variable that can be measured, either physically or chemically, or as a sequence of nerve impulses having a demonstrable frequency, duration and pattern The arrows do not have a mathematical significance here, they simply mean, "is accompanied by," "leads to," "causes," or "brings about"—that is, they denote a certain logical relationship in time Thus, a decrease in the cardiac output is accompanied by (or causes) a decrease in stretching of aortic walls, and this in turn leads to (is accompanied by or brings about) a decrease in the output of stretch detectors in the aortic wall and carotid sinus In these diagrams the nature of the central computational processes is not specified They are represented by Sherrington's term, *integration*

The similarities in these diagrams arise from a few basic principles of neural and endocrine physiology, as follows

1) The nervous system controls contraction of all smooth and striated muscles, and the secretion of exocrine glands

2) Both the nervous and endocrine systems function to give control at a distance, e g , by way of a digital-like transmission, the nerve impulse

3) The computation processes of the nervous system are brought about through infinitely graded degrees of stimulation or inhibition, (although synaptic transmission per se is quantal, see Katz, 1971)

4) Higher controls tend to be more complex than lower ones

5) No one type of control is independent of all others Thus, the overall pattern of controls includes priorities and compromises,

6) The creation of a control loop by introduction of feedback from detectors may convert a control into a regulation

The endocrine control systems share many of these characteristics, including the property of infinitely graded degrees of stimulation This particular property is a function of the cell membrane Moreover, many otherwise endocrine systems have at some stage a neural link, and vice versa For example, many of the endocrine mechanisms can be traced backward towards the origin of the control signals, back to the level of either the posterior or anterior lobe of the pituitary gland, which themselves are under the control of the hypothalamus—a neural control for the posterior lobe, and a neuroendocrine control for the anterior (Harris, 1955) The neurons that maintain these controls are presumed to accept neural input from other regions of the brain and spinal cord, and

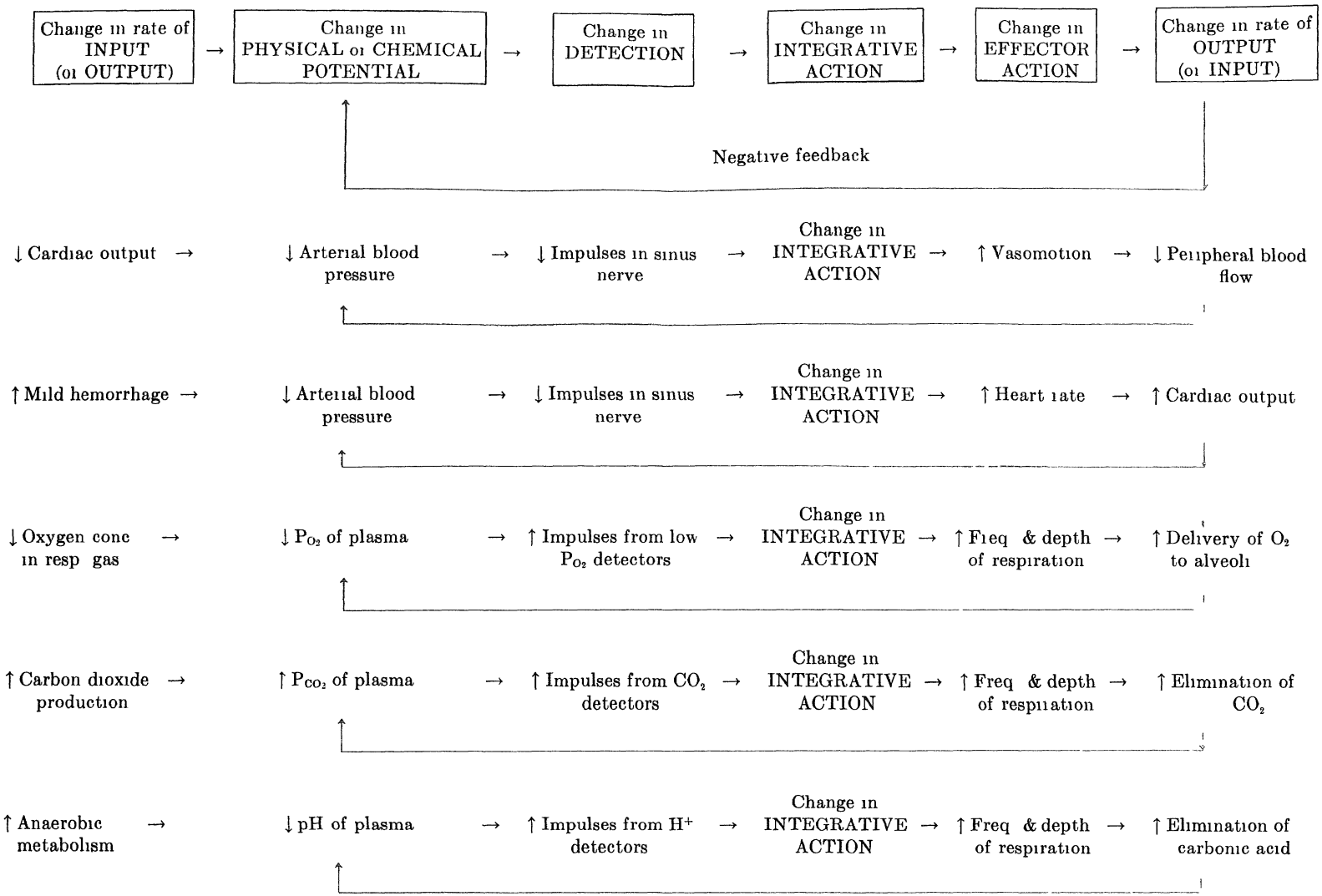

Change in rate of INPUT (or OUTPUT)
Change in PHYSICAL or CHEMICAL POTENTIAL
Change in DETECTION
Change in INTEGRATIVE ACTION
Change in EFFECTOR ACTION
Change in rate of OUTPUT (or INPUT)
Negative feedback
↓ Cardiac output
↓ Arterial blood pressure
↓ Impulses in sinus nerve
Change in INTEGRATIVE ACTION
↑ Vasomotion
↓ Peripheral blood flow
↑ Mild hemorrhage
↓ Arterial blood pressure
↓ Impulses in sinus nerve
Change in INTEGRATIVE ACTION
↑ Heart rate
↑ Cardiac output
↓ Oxygen conc in resp gas
↓ P_{O_2} of plasma
↑ Impulses from low P_{O_2} detectors
Change in INTEGRATIVE ACTION
↑ Freq & depth of respiration
↑ Delivery of O_2 to alveoli
↑ Carbon dioxide production
↑ P_{CO_2} of plasma
↑ Impulses from CO_2 detectors
Change in INTEGRATIVE ACTION
↑ Freq & depth of respiration
↑ Elimination of CO_2
↑ Anaerobic metabolism
↓ pH of plasma
↑ Impulses from H^+ detectors
Change in INTEGRATIVE ACTION
↑ Freq & depth of respiration
↑ Elimination of carbonic acid



Fig 9 103 Diagrams that illustrate how a change in the magnitude of a physiological variable becomes a change in potential that is detected by specialized cells They send signals to the nervous system as a result of which an effector response is created to alter the magnitude of the variable in an opposite direction This in turn alters the potential and hence the signal from detectors A simple feedback loop of this type will usually show oscillation

to convert this input into neural or endocrine messages transmitted to the lobes of the hypophysis (Fig 9 104) The hypothalamic neurons, therefore, are neuroendocrine transducers (Wurtman, 1970) They have also another interesting property, in that they are sensitive to the concentration of pituitary hormones, perhaps in a specific fashion appropriate to the nature of the system they control They may be sensitive also to more general changes, such as osmolarity, thermal gradients, or biochemical changes If their function is altered by, for example, the level of a circulating hormone, they may be designated as two-stage transducers, humoral-neural-humoral In the examples that at present are best known, the first of these stages is a negative feedback For example, ovarian steroids seem to act upon the hypothalamus to suppress neuronal activity, and hence the secretion of hypothalamic releasing substances, that ordinarily enhance secretion by the anterior pituitary gland

One other property of these systems should be noted The hormones of the endocrine system not only have specific actions upon their control systems as feedback loops, but they also have metabolic or other actions upon neurons in general This means that in the absence of a hormone, thyroxine, for example, many portions of the neural axis may show abnormal function Similarly, there is reason to believe that sex steroids have specific actions as negative feedback upon control neurons in the hypothalamus, but also more generalized effects along the neural mechanisms responsible for behavior These possibilities of interaction and control can be summarized by two statements, as follows

1) The central nervous system can exert a quick control of effector systems via neural messages, and also a slower and more prolonged control of certain organs by way of the endocrine glands

2) Similarly, the nervous system may gather information quickly from neural detectors, or in a slower and more prolonged fashion from the action of hormones or the influence of other types of generalized changes directly upon neuronal membranes and/or synaptic mechanisms

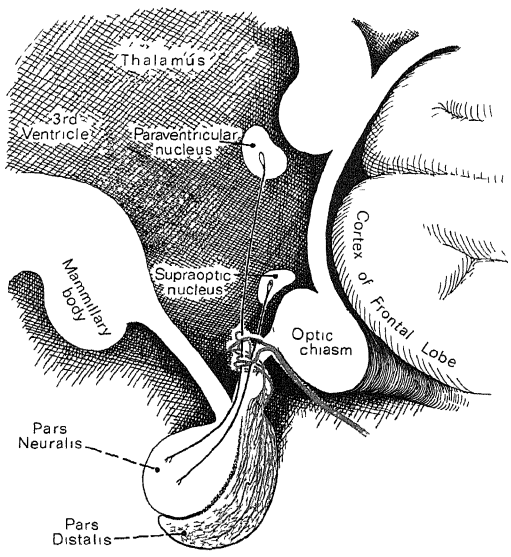


Fig 9 104 Drawing from photomicrographs of sagittal section through hypothalamus and pituitary gland where blood vessels have been injected with India ink Vessels that run along inferior surface of hypothalamus enter median eminence then contribute their blood to portal vessels distributed to pars distalis (anterior lobe) This portal system transports hypothalamic neurohumors to the pituitary gland

CIRCULATION

As Detweiler has indicated (sect 3), the quantities regulated in the circulation are arterial blood pressure, which is monitored as stretching of elements in walls of major arteries, and adequacy of blood supply in certain peripheral tissues, monitored as CO_2 tension (and pH) in, for example, skeletal muscle The latter, the constancy of local blood supply, is maintained almost entirely through local, presumably non-neural mechanisms that alter smooth muscle tone of small blood vessels in response to accumulation of carbon dioxide The general arterial blood pressure, by contrast, is preserved through a variety of mechanisms that control the following *a*) the force and frequency of the heart beat, and thus the cardiac output, *b*) the tone of vessel walls, both in arteries and veins, *c*) the degree of opening of arterioles through which blood flows out of the arterial system (this is the "peripheral resistance"), and *d*) mechanical forces such as the action of gravity, the contraction of skeletal muscles, and the magnitude of intrathoracic pressure, that affect the return of blood to the heart These mechanisms are controlled in large part by the autonomic nervous system, with participation of the somatic system and skeletal muscle, as in exercise, for

example, or in changes of posture There is interaction of these controls with those of other systems Thus, blood gas partial pressures affect the circulation as well as pulmonary ventilation by way of detectors in the carotid and aortic bodies

As noted in an earlier chapter, these peripheral mechanisms can be stimulated artificially in the medulla oblongata This is the level where the several reflexes are processed in regions known, respectively, as the cardiomotor, vasopressor, and vasodepressor mechanisms or "centers" They are resident in the reticular formation Similar stimulation evokes circulatory changes from hypothalamus, and from prefrontal or orbital areas of the cerebral cortex (p 9–127) It is the hypothalamic localization that is believed to be responsible for circulatory changes during exposure to heat or cold, and for those associated with emotional expression in decorticate animals The hypothalamus is also regarded as having an important role in the changes that occur in circulation during exercise (Folkow and Rubinstein, 1965) The higher levels including the limbic cortex (p 9-156) perhaps interact with the hypothalamus during emotional experience They may well serve also as components of feedback loops when exercise is initiated by cortical mechanisms One sees, therefore, that the circulation is controlled by a complex of mechanisms distributed almost the full length of the neural axis (Korner, 1971)

RESPIRATION

Control of pulmonary ventilation depends upon the sensitivity of peripheral and central detectors to partial pressures of carbon dioxide and oxygen, and is effected through changes in respiratory minute volume—that is, changes in respiratory frequency and depth (von Euler et al, 1970) (Fig 9 103) Any illusion that the system is a simple one begins to disappear, however, when one moves from consideration of peripheral to central mechanisms The nature of the central interaction of chemodetectors and the pulmonary stretch receptors is not fully understood, although it is known to take place in the medullary reticular formation, presumably upon the membranes of inspiratory neurons Moreover, the problem of how the respiratory rhythm is generated in the first place is unsolved As already noted, one possibility is that it represents a signal generator function of central neurons, in which a train of impulses is modulated so as to create a second order rhythm by way of feedback loops (chap 1, p 9–18) Control of pulmonary ventilation also interacts with other controls, for example, panting is a respiratory response to thermal stress Other interrelations are apparent during speaking, singing, or playing of a wind instrument, as well as with exercise and in sleep

Another characteristic of the control of pulmonary ventilation is the powerful behavioral reactions that can be evoked as part of it Consider, for example, a subject with his nostrils and mouth forcibly covered, or a subject held involuntarily under water, or in the presence of a gas such as ammonia, formaldehyde, or HCl The behavioral component of the regulation may supersede all other kinds of behavior The significant point is that regulations are not only automatic, involuntary, and reflex They may command the highest levels of the nervous system and include a great variety of patterns of behavior In the case of pulmonary ventilation, whereas the peripheral mechanisms and the reflex connections are reasonably well known, and the properties and location of the brainstem neurons can be described in some detail, the genesis of the behavioral reactions is all but unknown

WATER BALANCE

Hypothalamic mechanisms control both the rate of water loss through the kidneys and the rate of water ingestion in drinking Water loss was the first of these to be analyzed experimentally (sect 5, chap 1) A disorder of this control has been known for centuries, the condition named diabetes insipidus, where a patient involuntarily excretes a large volume of dilute (i e, "insipid") urine, and in compensation drinks a large volume of water daily Early experiments did not disclose the nature of the diabetes, because it was assumed that the effective lesion would prove to lie either in the hypothalamus or in the pituitary gland Later experiments showed that both structures are involved (Fisher et al, 1938), because secretion of the antidiuretic hormone in the posterior lobe is controlled by neurons of the supraoptic nucleus (sect 7, chap 3) By perfusing the base of the brain with blood having an osmotic pressure either higher or lower than normal, Verney (1947) and others demon-

strated that this region contains a mechanism sensitive to water concentration in the extracellular fluid Whether the detector is, in fact, neurons of the supraoptic nucleus, or whether it lies near these neurons, or whether it is elsewhere and projects to these neurons is not entirely clear Cross and Green (1959), as well as Brooks and his associates (Koizumi, et al, 1964), have recorded changes in single unit activity from this nucleus in correlation with changes in osmolarity (Figs 9 105 and 9 106) Moreover, cats with the hypothalamus removed except for an "island" that includes the supraopticohypophysial system do not exhibit diabetes insipidus (Woods and Bard, 1960) In any event, for control of water loss via the kidney this system is one that has two stages of transduction First, there is transduction of water concentration ("activity") into nerve impulses, second, there is transduction of this neural activity into secretion of the antidiuretic hormone (Dreifuss et al, 1971)

A second kind of transduction evidently occurs within the major blood vessels, especially the right atrium, where changes in hydrostatic pressure bring about changes in secretion of antidiuretic hormone (Henry, et al, 1956, Share, 1968) By these two systems the rate of water loss is made to vary in response to changes in the concentration of water in the extracellular fluid and the volume of this fluid as reflected in changes in central venous pressure

Water Intake. For control of water intake the system contains additional detector mechanisms and reflexes The first of these begins with perception of dryness in membranes of mouth and throat, as suggested by Cannon (1918) and later more fully described by Epstein and his associates (1964) The latter have found in rats that interruption of salivary ducts and extirpation of salivary glands leads to a condition where the animals must necessarily drink when they eat dry food, even though their general bodily state may not require water A second mechanism is the water concentration in fluids of the body, and a third originates in the pressure-sensitive neurons of the great veins, as in the control of secretion of the antidiuretic hormone The fourth and last is a mechanism recently discovered by Fitzsimons (1969) He observed that the amount of water drunk by nephrectomized rats is less than that taken by animals with occluded ureters, when conditions that induce drinking are imposed upon the animals From this he concluded that the kidney is placing in the circulation something that facilitates drinking The material that does this is renin, the enzyme that produces angiotensin II The most recent experiments have shown that angiotensin II induces drinking when it is perfused directly into the hypothalamus (Fig 9 107)

Lesions in the lateral hypothalamus abolish drinking (Montemurro and Stevenson, 1955 and 1956), except that arising from dryness of the mouth (Epstein and Teitelbaum, 1964) The latter can be interrupted by section of appropriate cranial nerves The ingestion of water is a more complicated form of behavior than those normally responsible for ventilation and other similar regulations It may involve really heroic effort for securing of water (Wolf, 1958), or it may be as simple as the turning of an animal's head to a water spout for a brief period of lapping At some level in the neural

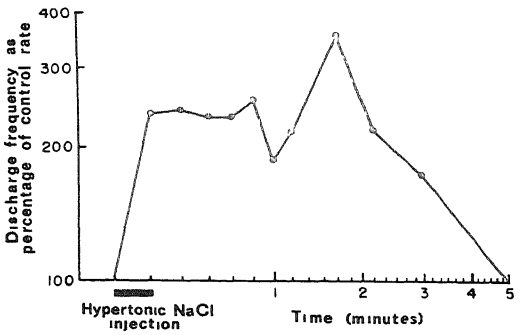


Fig 9 106 Increased frequency of discharge of neurons of supraoptic nucleus (*ordinate*) during and after intracarotid injection of hypertonic sodium chloride solution (0 5 ml of 1 M NaCl solution) (Koizumi et al 1964)

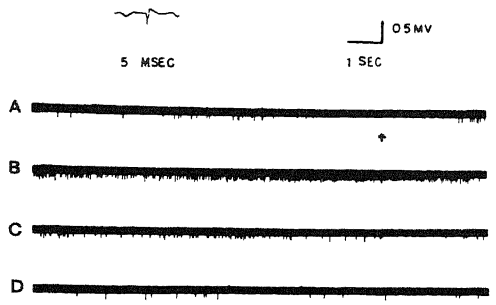


Fig 9 105 Action potentials from single neurons in supraoptic nucleus *A*, before injection of hypertonic NaCl solution *B* 30 sec after end of injection *C* 1 min after *D* 3 min after (Koizumi et al 1964)

or neuroendocrine circuits some completely unknown mechanism creates an awareness of need for water At any instant, for example, any adult subject can state whether or not he feels thirsty This awareness is presumed to require cortical function, but there is no evidence as to the nature or locus of this highest level component In the hierarchy of mechanisms for regulation, water balance is somewhere near the middle The body cannot sacrifice either circulation or respiration as a device for preserving water balance, yet water is more vital to the body than temperature regulation, energy balance, or reproduction

ENERGY BALANCE

Still more complicated than the regulations heretofore described is the preservation of energy balance This is assumed to be a regulation because of the constancy of its end product—body substance as measured by the body weight of an adult subject Mechanisms of gain of energy are known, in higher animals they are solely by way of food intake Mechanisms of energy loss are more diverse Energy can be lost as heat, or as work in a truly physical sense—i e , in raising a weight or giving it a momentum Or energy can be stored within the body in the form of protein during growth or recovery from starvation, as carbohydrate (glycogen) during recovery from a brief fast, or as fat in almost unlimited quantities Because each of these categories of processes has been studied for many years, a large amount of information is available about all of them This

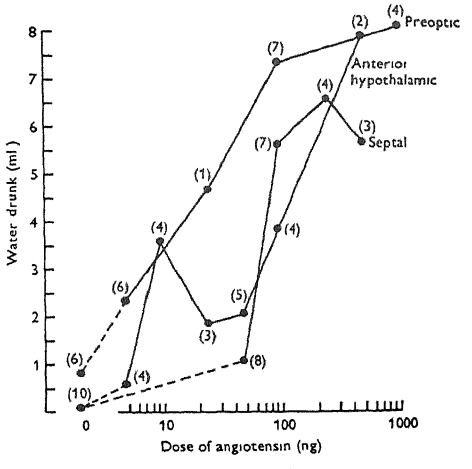


Fig 9 107 Water drunk (*ordinate*) by three rats as a function of amount of angiotensin (*abscissa*) injected intracerebrally (Epstein et al 1970)

BODY WEIGHTS, FOOD INTAKES AND FINAL BODY FAT CONTENTS OF PREVIOUSLY FORCE FED RATS ALLOWED ACCESS TO FOOD

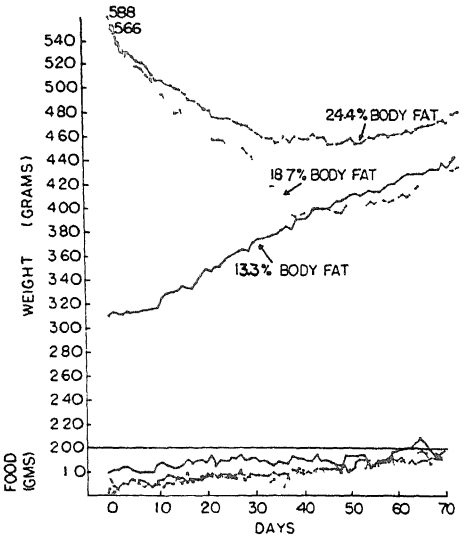


Fig 9 108 Two rats (566 588) that had been made obese by forced feeding ate a reduced amount of food when they were fed ad libitum until their body weight and estimated body fat concentration returned to the normal levels (Cohn and Joseph 1962)

is not the kind of information needed, however, to account for preservation of energy balance. What is needed is a knowledge of how the body—with variable levels of work output, under variable conditions of environmental temperature and humidity, yet at a given age and growth potential—how the body under these circumstances manages to adjust food intake so that body weight changes slightly or not at all

The possibility that there is some overall system for this regulation is suggested by a few but convincing observations First, there is the common experience of adult men who find that their body weight is almost constant in spite of variations in activity, and often in the face of efforts to gain or lose Second, experiments by Cohn and Joseph (1959) proved that animals made obese by overfeeding would not eat a normal quantity of food until their body weight had returned to the control level (Fig 9 108) Third, lesions in the hypothalamus may cause a failure of feeding (Anand and Brobeck, 1951), or hyperphagia and obesity (Tepperman et al , 1941) Fourth, injury to the hypothalamus will profoundly alter regulation of body temperature, and thus the energy converted into heat and lost from the body (see later in

this chapter) Fifth, there is evidence that injury to the hypothalamus or the brain just ahead of it will either increase or decrease motor output as locomotion

What has been said about the results of hypothalamic injury can be said in most instances in reverse about hypothalamic stimulation For example, if a lateral hypothalamic lesion abolishes feeding, stimulation of the lateral hypothalamus will induce feeding (Leibowitz, 1971) Again, if locomotor activity is abolished by lateral hypothalamic lesions, it may be induced by electrical stimulation of this part of the brain

How the hypothalamus, either independently or in collaboration with the limbic system, rhinencephalon, or other parts of the brain, integrates all of the processes required for energy balance is unknown At least three kinds of "sensation" are created—hunger-satiety, warmth-cold, and tired-rested Receptors for the warm-cold pair are found in the skin and oronasal membranes, as well as in the rostral hypothalamus or preoptic region Where the other kinds of detection occur is not certainly known With reference to the hunger-satiety state, the possibilities under consideration include the gut, liver, and hypothalamus (see sect 2, chap 1)

Regulation of Body Temperature

The temperature of many warm-blooded (*homoiothermic*) animals remains practically unchanged although the surrounding temperature may vary between 0° and 50° C or upward On the contrary, the body temperature of a cold-blooded (*poikilothermic*) animal such as a frog, turtle, etc , is practically that of its environment during laboratory tests (Fig 9 109) In natural conditions these latter animals preserve an almost constant temperature by behavioral responses that include migration and hibernation

The normal human body temperature recorded from the mouth is usually given as 37° C, with the rectal temperature one degree higher The figure varies between individuals, ranging from 35 8 to 37 8° C orally Variations also occur in any one individual throughout the day—a difference of 1 0° or even 2 0° C occurring between the maximum in the late afternoon or early evening, and the minimum between 3 and 5 o'clock in the morning Temperature of the several regions of the body depends upon whether they are major sources of metabolic heat (e g , liver, heart), or locations where heat is given off (e g , skin, upper respiratory tract) (Fig 9 110) Strenuous muscular exercise causes a temporary rise in body temperature that is proportional to the severity of the exercise, the level may go as high as 40 0° C

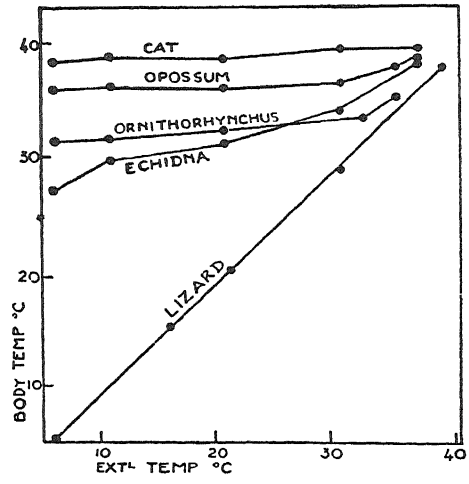


Fig 9 109 Body temperature of five species of animals as a function of environmental temperature Duration of each exposure was 2 hr (Martin 1930)

Stability of body temperature varies with the size of the body In small animals or in human infants where metabolic rate is relatively more intense and surface area relatively greater, the oral and even the rectal temperature may vary through a range perhaps twice that of larger animals or human adults This reflects certain principles of the physics of heat storage and loss, rather than any inherent difference in the regulating system In a small body the amount of heat "stored" is a smaller fraction of the metabolic rate than in larger individuals In the latter the stored heat provides an "inertia" that tends to counteract thermal changes Heat storage is a function of thermal capacity, which varies with body mass, heat loss, by contrast, varies generally with surface area The most lucid account of these and other aspects of thermal exchange is the one in Kleiber's monograph (1961)

Under basal conditions when no external work is being done, all the metabolic energy ultimately appears as heat, and amounts to approximately 1 kcal /kg of body weight/hr

Since the body is composed chiefly of water, and has a specific heat of a little less than one, the body temperature would rise about 1° C/hr if no heat were lost externally With strenuous physical exertion more than three-quarters of the increased metabolism appears as heat within the body, the remainder is converted either to work or to heat in the external system Fever increases the metabolic rate by some 13% for each degree rise in mean body temperature (C) This change in metabolism corresponds to a temperature coefficient (Q_{10}) of between 2 0 and 3 0 overall

The body's heat-regulating mechanisms are supposed to be in abeyance below an internal temperature of about 23° C, because the underlying neural mechanisms are suppressed The body then gains or loses heat like an inanimate object The lower lethal temperature for the human is about 26° C, although lower temperatures have been induced with survival during therapeutic hypothermia Death is due to cardiac failure The upper lethal limit is not clearly defined, cases are recorded to have recovered after attaining a rectal temperature of 43 5° C Probably the average upper limit is about 43° C

HEAT LOSS

Heat is lost from the body through *a*) radiation, conduction, and convection from the skin, *b*) warming and humidifying of inspired air, *c*) evaporation of sweat and insensible perspiration, and *d*) via urine and feces Of these several routes only *a*) and *c*) are directly under physiological control Radiation is responsible for about 50% of the total heat loss and convection for about 15% (Table 9 6) Urine and feces account for only about 2% or less of the total heat loss Most of the remainder (about 30%) is lost in the evaporation of water These percentages are given for an average man doing light work with an overall energy exchange of about 3,000 kcal /day

Radiation. The rate of cooling of any object varies with the thermal gradient between its surface and objects in its environment This gradient can be altered by physiological mechanisms that change skin temperature These mechanisms include the following adjustments in the blood-vascular system

a) *Redistribution of Blood* By redistribution of blood flow the temperature of the skin may be adjusted from any temperature from about

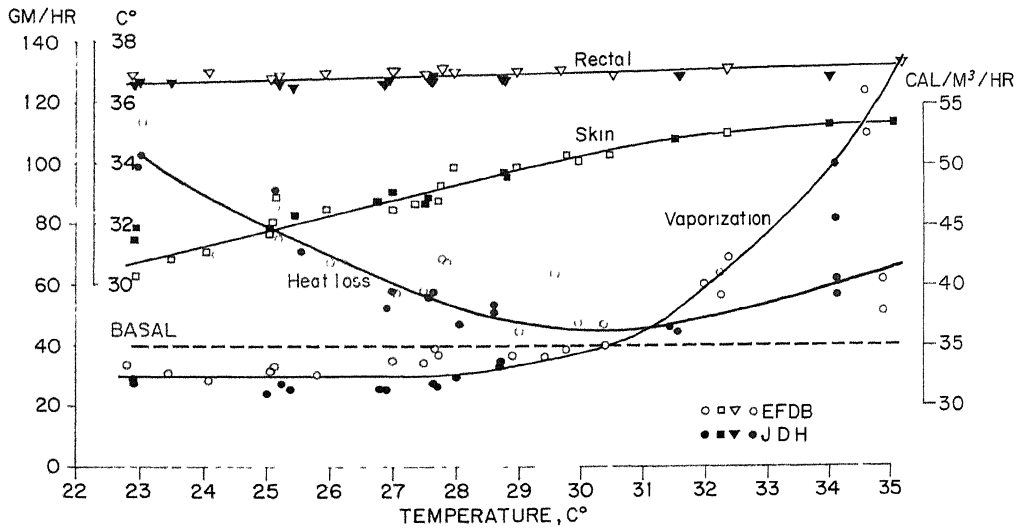


Fig 9 110. Graphs showing the influence of environmental temperature upon rectal and skin temperature (with mean body temperature derived arithmetically from them) Metabolic rate rises during cold exposure mainly as a result of shivering Vaporization rises in the heat mainly because of sweating Crosbie et al 1961 replotted

15° C (during cold exposure) up to the central body temperature At ambient temperature of 34° C the quantity of blood circulating through the skin may be as great as 12% of the cardiac output (DuBois, 1939) Scott, et al (1940) found that blood flow to the fingers could increase from 1 ml /min /100 gm of tissue to a figure of 80 or 90 ml /min upon exposure to heat Fingers, therefore, serve as highly efficient radiators of heat

b) Variations in Blood Volume An unacclimatized man is capable of increasing total circulating blood volume by about 10% in 2 to 4 hr on exposure to heat of severity sufficient to cause a diffuse cutaneous vasodilatation The blood is diluted by fluid drawn into the circulation from the tissues, chiefly the skin, muscles and liver, cells are added from the spleen Hematocrit drops and concentration of plasma protein decreases The reverse changes occur during exposure to cold A sudden transition from a warm to a cool environment causes a rapid decrease of blood volume associated with a marked diuresis

c) Increased Circulation Rate An increased cardiac output ensures a rapid blood flow through the dilated cutaneous vessels This increase and the enhanced blood volume promote the transportation of heat from the interior to the surfaces of the body and so enhance the loss of extra heat

With reference to radiation of heat or infrared radiation, the human skin of whatever "color" functions as a nearly perfect "black body" in the physical sense (Hardy, 1949) In a standing posture about 85% of the total skin area is exposed for heat radiation, the remainder being the opposed surfaces of, e g , axillae and other similar regions of the body The percentage exposed is increased during heat stress and decreased in the cold by postural changes Clothing modifies radiant heat loss by altering the temperature of the surface where the heat exchange is occurring

Convection The most important factor influencing heat loss by convection is air movement For any given air temperature the loss increases with the square of the wind velocity up to 60 m p h , beyond this velocity there is little further increase

Evaporation of Water. It is obvious that the nearer the temperature of the environment comes to that of the blood the smaller will be the amount of heat lost by radiation and convection Above an ambient temperature of 38° C the body gains heat by radiation from the environment This heat is then dissipated, together with heat produced within the body, by processes in which water is converted from a liquid to a vapor state at surfaces of the body The heat absorbed in evaporation of 1 ml of water amounts to 0 58 kcal Even at ordinary room temperatures when there is no obvious perspiration, the heat lost through evaporation from the lungs and skin amounts to about 17 kcal /hr This is approximately one-fourth of the basal heat production About two-thirds of this is lost from the skin as insensible perspiration, the remainder from respiratory passages At higher temperatures the proportion of heat lost by vaporization of water, mainly from the skin, increases dramatically, so that at ambient temperatures above 35° C it accounts for nearly all of the heat lost (Fig 9 110)

In animals that pant a comparable increase in vaporization occurs via the respiratory system Rats enhance their heat loss by spreading saliva upon their fur (Hainsworth, 1967) The rate of evaporation of this water is influenced inversely by the degree to which the atmosphere is already saturated with water—by its relative humidity This is why sweat drips instead of evaporating when the ambient air is humid A man can maintain a normal temperature in an atmosphere of over 100° C provided the air is perfectly dry On the other hand, in a damp atmosphere a temperature of 50° C causes body temperature to rise rapidly

Sweating. Sweat is a weak solution of so-

Table 9 6
Partitioning of heat loss of a subject whose daily energy exchange is 3 000 kcal equivalent to a daily schedule that includes light work

	Kcal	Per Cent
a) Radiation, convection and conduction	1950	65
b) Evaporation of water from skin and lungs, and liberation of CO_2	900	30
c) Warming inspired air	90	3
d) Urine and feces (i e , heat of these excreta over that of the food and water)	60	2
Total daily heat loss	3000	100

dium chloride in water, together with urea and small quantities of potassium ion, other electrolytes, and lactic acid It has a specific gravity of from 1 002 to 1 003, and a pH varying from 4 2 to 7 5 The concentration of NaCl varies from 50 to 100 mEq /liter When sweating is profuse the concentrations of electrolytes rise towards the higher of these values, whereas with acclimatization the percentage is said to decline Nevertheless, even after acclimatization the performance of strenuous exercise, e g , marching, may lead to electrolyte and water depletion If the water is replaced without salt, cramps occur in muscles of limbs and abdominal wall They can be relieved by administration of NaCl tablets, or by assimilation of the electrolyte that is found naturally in most foods

Control of Sweat Secretion The sweat glands, which number over 2,500,000 in a man living in a temperate climate, are cholinergic although their innervation is by means of postganglionic fibers of the thoracolumbar division of the visceral motor system Sweating is controlled by mechanisms that are present in almost all levels of the nervous system For example, in the initial stages of muscular exercise sweating is apparently initiated by the discharge of impulses from the motor cortex, inasmuch as it occurs before there is any significant increase in heat load Later on in exercise a rise in hypothalamic temperature acts upon detectors at the hypothalamic level (Fig 9 111) Spinal centers exist for segmental control of sweating, this can be demonstrated in quadriplegic patients who sweat reflexly in parts of the body innervated below the level of the transverse lesion of the spinal cord

The few observations that have been made upon the secretion pressure of sweat indicate that it is high, 250 mm of Hg or more Sweat is therefore a true secretion and not simply a filtrate The rate of sweating may be really surprising, it may be as high as 1 6 liters an hour If it were all evaporated this would remove over 900 kcal of heat from the body in each hour

HEAT PRODUCTION

When the environmental temperature is in the range between 28° and 31° C, the basal heat production can be dissipated to the environment by radiation, convection, and insensible vaporization of water by an unclothed male subject If the environmental temperature is lowered below these levels, loss by radiation and convection become progressively greater, and the mean body temperature falls because the periphery becomes cooler At a level of about 23° C (sometimes called the "critical temperature" for a nude subject) shivering begins (Fig 9 112), and intensifies the rise in heat production that has occurred imperceptibly as the temperature fell below 28° C (Fig 9 110) This rise in metabolic rate occurs principally in skeletal muscles as a result of increased tension, even before shivering is initiated When shivering is intense the overall heat production may be as much as three times greater than the basal rate It is evident that this is a mechanism that is controlled by the somatic division of the nervous system, but that subserves an unconscious and automatic function in regulation of body temperature After the skeletal muscles are paralyzed by a drug such as curare, an animal loses the power to maintain a normal temperature in a cold environment Its ability to dissipate extra heat is not impaired, however, so that it can still resist high temperatures

CENTRAL MECHANISMS FOR TEMPERATURE REGULATION

Integration of the autonomic mechanisms that control heat loss and the somatic mechanisms that govern heat production, posture, and behavior, is accomplished in laboratory animals in the hypothalamus (In man the behavioral adjustments to thermal stress presumably involve higher levels of the nervous system) The significance of the hypothalamus can be illustrated by review of three types of evidence, as follows

a) Injury to the hypothalamus almost invariably leads to changes in body temperature If experimental lesions are placed in the rostral portion, the animal may exhibit a severe fever that may progress to death This fever is regarded as a release phenomenon analogous to decerebrate rigidity The rostral hypothalamus apparently contains an inhibitory mechanism that normally acts to suppress the mechanisms of heat production, and that also facilitates the stimulation of mechanisms of heat loss These particular results would follow the interruption of the loop by which information is returned to

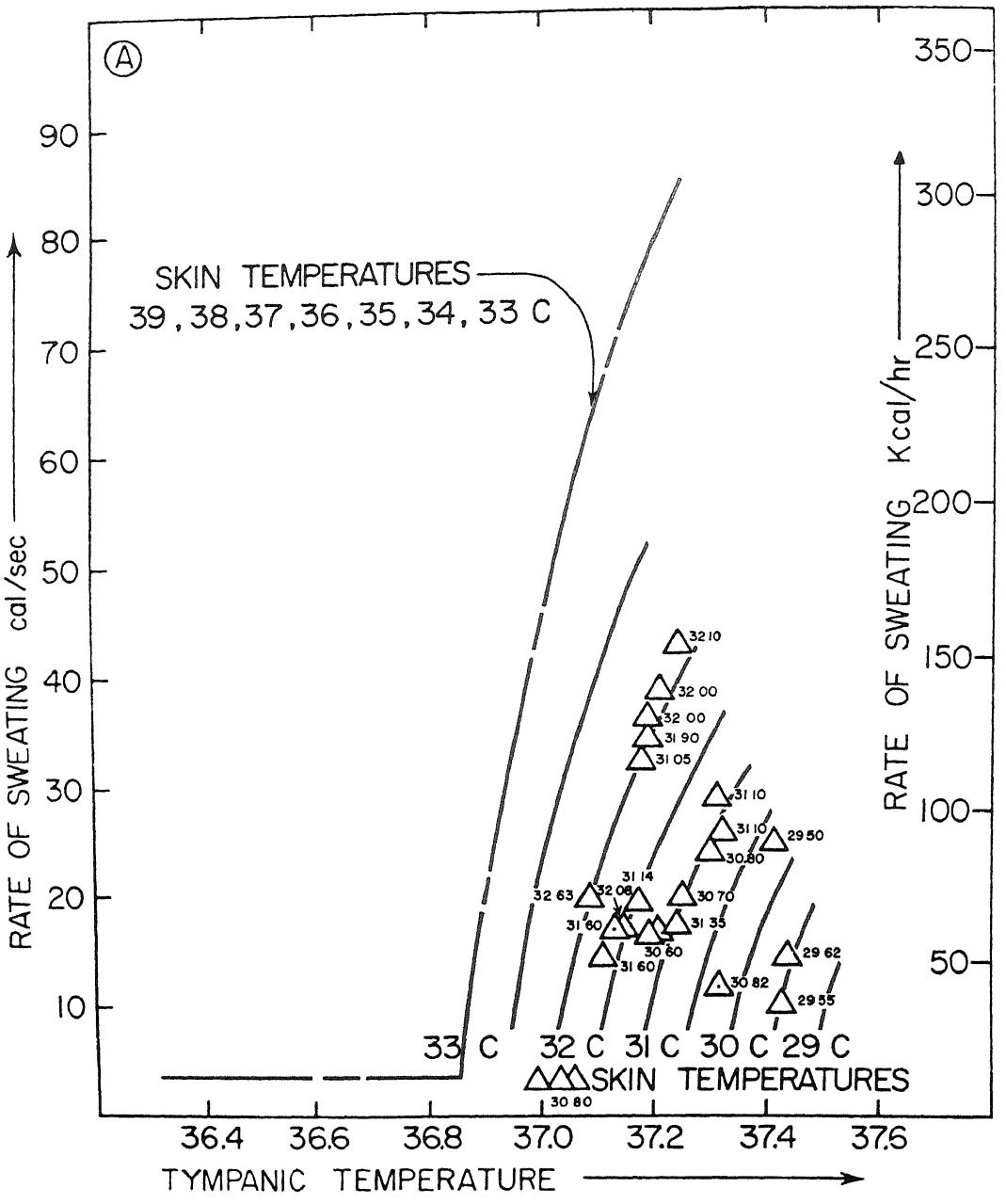


Fig 9 111 Rate of sweating of human subjects having differing central (*tympanic*) temperatures (*abscissa*), and a variety of skin temperatures (the family of *curves*) (Benzinger 1969)

the controller with a negative sign The usual therapy for fever of this type, particularly when it occurs in human patients after neurosurgical operations, is forced cooling by means of ice packs and cold water enemas A more specific therapy is administration of a barbiturate, Ranson and his associates (Beaton et al , 1943) discovered that such a drug selectively decreases the activity of fever-producing mechanisms and so permits body temperature to return to normal If the hypothalamic lesion is more caudal, and especially if it extends into the lateral hypothalamic areas, the patient or experimental animal will have a low

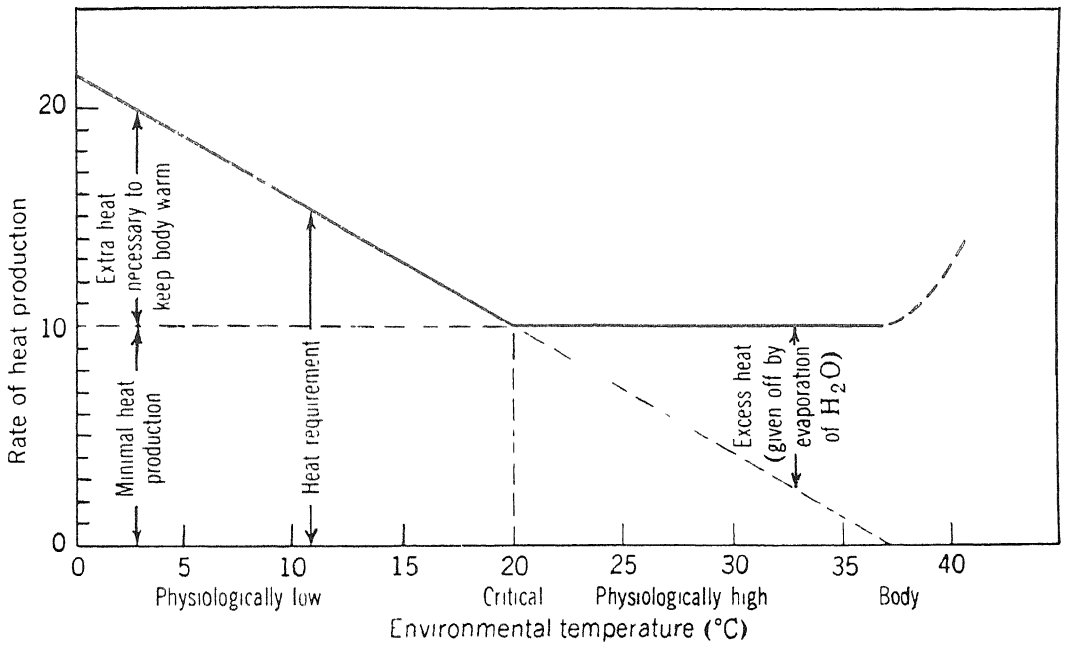


Fig 9 112 Diagram that illustrates the influence of environmental temperature upon metabolic rate At 20°C on this graph the subject can lose all metabolic heat without extra evaporation of water similarly body temperature can be maintained without shivering (Kleiber 1961)

temperature, particularly if the environment is cool This hypothermia can be corrected by artificial heating, but if it is not so treated it, too, may terminate fatally

b) The second type of evidence came from experiments such as those of Karplus and Kreidl (1910), and of Ranson (1936 to 1937) and his associates, who discovered that most of the reactions needed for temperature regulation can be brought into play by electrical stimulation of the hypothalamus The major exception to this statement is shivering, which cannot be induced by any conventional type of electrical stimulus It is a nonsynchronized but rhythmic phenomenon that requires a pattern of activation different from the varieties physiologists employ

c) The most explicit type of evidence comes from experiments in which the hypothalamus has been artificially warmed or cooled. If the temperature of the hypothalamus is changed by high frequency warming through electrodes, or by conduction via thermodes, an artificial feedback loop is created that is open, or independent of what is taking place in the control system Experiments of this variety were first done by Barbour (1912), who passed warmed or cooled water through fine tubes implanted in the brain of rabbits Later the technique was refined by Magoun et al (1939), who oriented electrodes using a stereotaxic method, for passage of a "diathermy" type of high frequency current through the hypothalamus The warming of a sensitive region that lies over the chiasm in cats, and ahead of it in monkeys, led to polypnea, panting, and sweating on foot pads Other investigators have extended the observations by use of implanted thermodes in unanesthetized animals, and have shown that this region is sensitive to cooling as well as heating It apparently serves as a thermostat for the rest of the body—as a thermal detector that originates a negative feedback (Benzinger, 1969) In a sense, it utilizes the rest of the body to stabilize its own temperature, cooling the body when it becomes warm, and retaining heat in the body and stimulating heat production when it is chilled (Fig 9 113) Investigators using microelectrodes have studied the function of single units in this region Although some neurons seem to be insensitive to temperature, and have been proposed as generators of a constant or set point signal, others

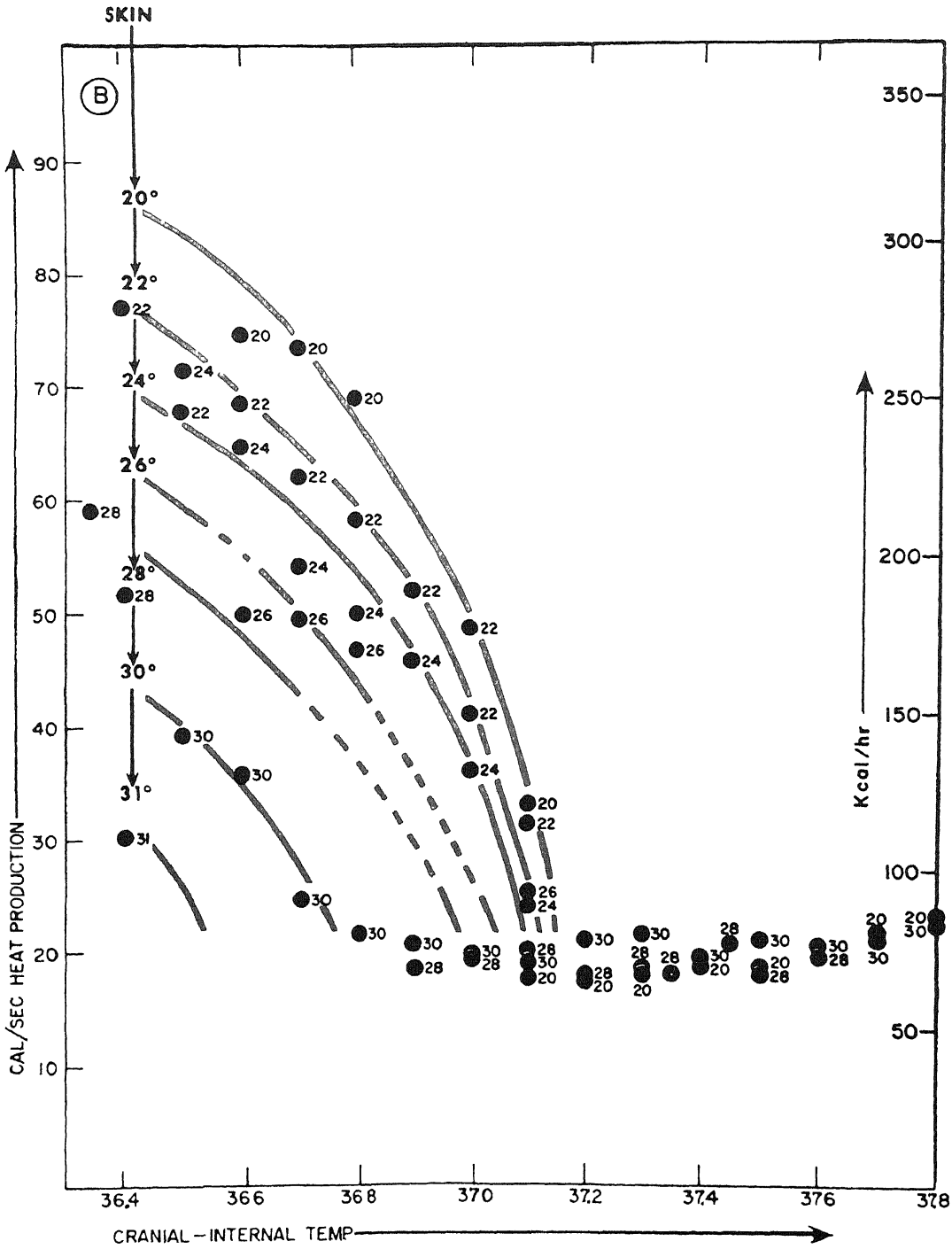


Fig 9 113 Relationship between central (*tympanic*) temperature and rate of heat production with skin temperature as another variable (family of *curves*) (Benzinger 1969)

respond to heating whereas still others are stimulated by cold (Figs 9 114 and 9 115)

In addition to this central detector and feedback loop, there is another similar loop that arises from thermal detectors in the skin (Figs 9 111 and 9 113) Its function is illustrated in Figure 9 116, where the response of an unanesthetized dog to heating of the hypothalamus is

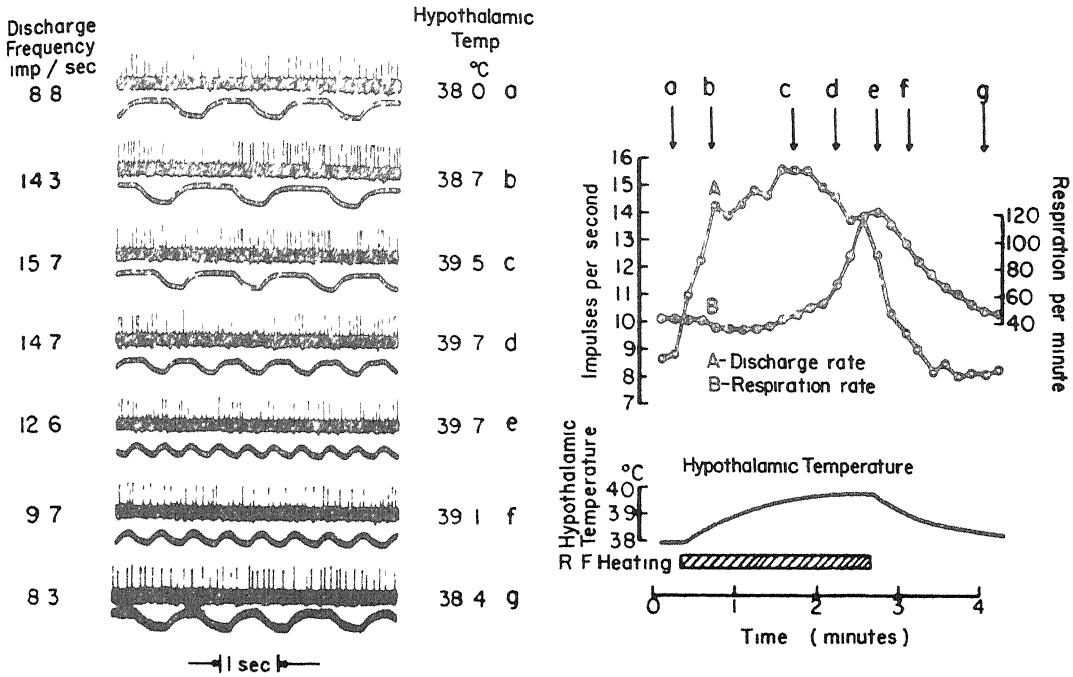


Fig 9 114 Action potentials recorded from single units in rostral hypothalamus during experiment in which the hypothalamus was warmed artificially through an implanted electrode The graphs indicate the course of the experiment and the frequency of firing of the unit (Nakayama et al 1961)

shown to depend upon ambient temperature Stolwijk and Hardy (1966) have devised a mathematical formula that combines these two kinds of signals According to their concept the body does not respond to a change in skin temperature when there is no central change, and likewise does not respond to central change in the absence of peripheral deviation They have concluded, therefore, that the central "integration" is some function of a mathematical product of the deviations in temperature in the two locations

The nature of the set point of this system has been considered by both Hardy (1965) and Hammel (1965), the latter suggested that the elevated temperature of exercise represents a change in set point that originates in neural adjustments inducing or accompanying the exercise A similar explanation may account for the elevated temperature of infectious fevers, but on a biochemical basis at the neuronal level (Eisenman, 1969). Toxins known as *pyrogens* are liberated by bacteria and can be isolated also from white blood cells (Wood, see Moore et al , 1970) In effect they stabilize the membrane of the central thermodetector neurons so that a higher temperature is required

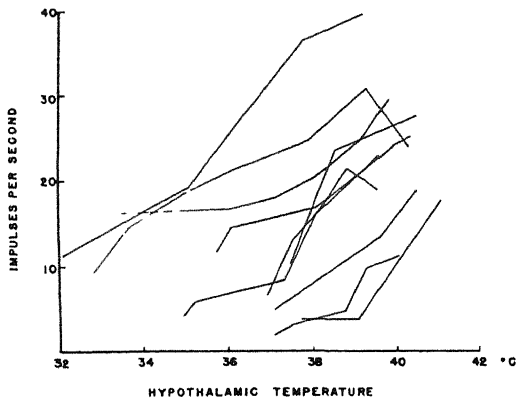


Fig 9 115 Frequency of firing of warm-sensitive units in rostral hypothalamus as a function of hypothalamic temperature (Hardy 1965)

to induce the negative feedback The possibility that the chemical intermediary of the action of pyrogens is the class of compounds known as prostaglandins has been suggested recently by Vane (1971). His suggestion is based upon the discovery that aspirin, a drug with pronounced antipyretic action, inhibits the synthesis of prostaglandins, and upon reports that prostaglandins injected directly into

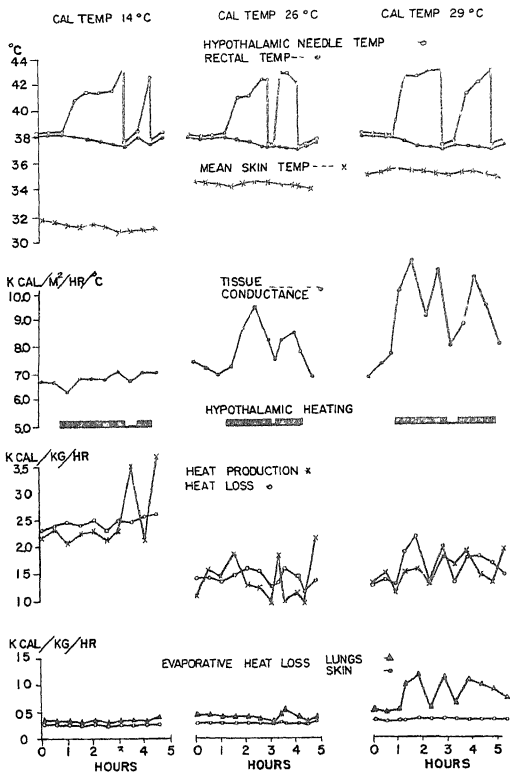


Fig 9 116 Records of experiments in which hypothalamus of a dog was warmed when the animal was exposed to three different environmental temperatures 14° C 26° C and 29° C Tissue conductance is a measure of vasodilatation in skin vessels (Fusco et al 1961)

the hypothalamus induce fever Prostaglandins are widely distributed compounds whose biological significance is not entirely clear, they are synthesized in the body from polyunsaturated fatty acids, and are believed to be related to the cyclic adenosine monophosphate system (Horton, 1969)

ENDOCRINE GLANDS IN THERMOREGULATION

The thyroid and adrenal glands play significant roles in the regulation of body temperature The calorigenic effects of the secretions of these glands are well known (sect 7, chap 4) Cannon (1932) observed that exposure to cold causes an increase in the rate of the denervated heart of a cat, as a result of epinephrine secretion This hormone exerts a calorigenic effect which is immediate and of short duration A less immediate and much more prolonged increase in heat production is brought about by stimulation of the thyroid gland Rats exposed over a period of 3 weeks to low temperatures (7° to 12° C) show thyroid hyperplasia and a rise in metabolic rate of as much as 16% Thyroidectomized rats, on the contrary, show little rise in metabolic rate under the same conditions Uotila (1939) found that this stimulation of the thyroid did not occur if the pituitary stalk had been previously sectioned He thus provided the earliest evidence that secretion of the thyrotrophic hormone by the anterior pituitary gland is under the control of hypothalamic mechanisms This control is now known to be exerted by way of thyrotrophin releasing factor (sect 7, chap 2)

Some years ago Ring (1942) demonstrated another interesting mechanism of protection against cold exposure when he found that thyroid hormone potentiates the calorigenic action of epinephrine This means that thyroidectomized animals not only are deficient in the action of thyroid hormone on metabolic rate, but also that the epinephrine they secrete in the cold has a reduced potency in elevating heat production even in a transient way

8 / HIGHER NEURAL FUNCTIONS

Sleep

Although sleep is known to everyone from personal experience, it has a special significance in neurophysiology Consider, for example, the results of a neurological examination performed upon a sleeping subject If the subject remains asleep throughout the examination, the examiner will note what appear to be catastrophic deficiencies of central neural function—as contrasted to the waking state These include loss of muscle tone and changes in both cutaneous and deep reflexes, elevated threshold for all types of sensory stimulation, almost complete dissociation from the environment, visceral changes including fall in metabolic rate, lowering of central body temperature, perhaps an elevation of skin temperature, and a lowering of blood pressure and heart rate, and a respiratory pattern different from that of the waking state to a degree that in infants may include periodic breathing This constellation of neurological signs would be consistent with some profound neurological injury, were it not that they disappear upon waking

The transition from sleep to waking, and the preservation of the waking state are functions of a mechanism that lies in the hypothalamus and the tegmentum of mesencephalon and pons It can be brought into activity by sensory stimulation, as for example an alarm clock It is believed also to contain, however, circuits of neurons that give it what appears to be an inertia (in the physical sense)—a momentum so that it keeps an individual awake in the intervals between arrival of particular sensory stimuli It may actually include also neurons that can generate signals, as described in chapter 1

What serves as a switching mechanism to convert the central nervous system from sleep to waking or vice versa is not known

EEG OF SLEEP

The principal tool for study of these phenomena is the electroencephalogram (EEG), described in chapter 6, p 9–116, which exhibits different patterns in the waking and sleeping states (Fig 9 117) It is true that some of the patterns seen during sleep occur also in waking, and vice versa (Jouvet, 1967) Nevertheless, the correlation between the EEG record and sleep is quite good, for studies on experimental animals and infants there is no equally useful technique for analyzing sleep

A subject who is awake and alert to his environment shows in his EEG from all cortical areas a random pattern of fast waves of low voltage that is believed to arise from lack of synchronization of the neurons responsible for the electrical activity in the area under the electrode If the subject closes his eyes, the pattern changes and begins to show the α rhythm (Berger, 1929) If the subject then becomes drowsy the α waves become less prominent, and larger, slower waves begin to appear in bursts lasting a few seconds With the onset of light sleep the α waves are gone, the record shows mainly low voltage fast waves again, and sleep "spindles" are seen (Fig 9 118) These latter are short bursts of waves having a frequency of 14 to 16/sec , with a gradual increase and then a decrease in voltage so that the overall profile of the complex is something like an elongated, slim American football During deep sleep these spindles likewise become less prominent and may even disappear, the record is then dominated by slow waves of relatively high voltage that may have been noted first during the stage of light sleep These waves mark the condition known as "slow sleep," more correctly, sleep characterized by synchronized cortical activity When wakening takes place, the pattern returns very quickly to the asynchronized, fast, low voltage activity and the process of *activation* of the EEG is said to have occurred

During the stage of sleep with slow waves (synchronized cortical electrical activity) there may occur every now and again another kind of record that ressembles the activated pattern Waves are momentarily fast and of low voltage (Derbyshire et al , 1936, Rheinberger

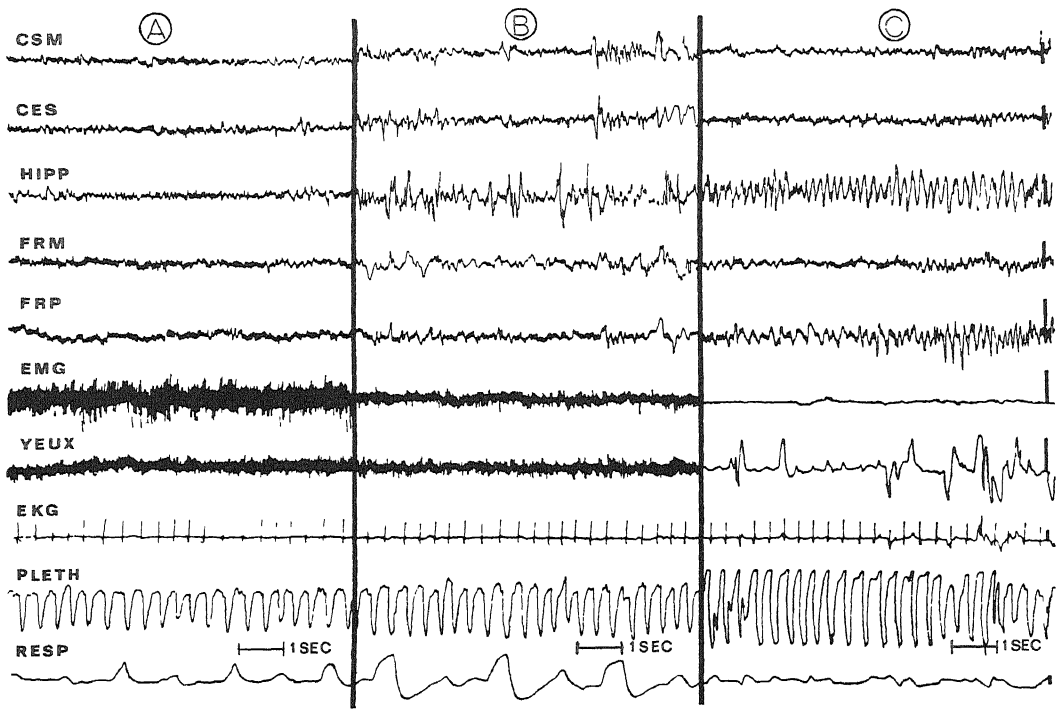


Fig 9 117 Electrical recordings that show differences in (*A*) waking state (*B*) slow wave sleep and (*C*) paradoxical sleep (Jouvet 1967)

CSM cortex sensorimotor	FRM retic form mesen
CES cortex ectosylvian	FRP retic form pontile
HIPP ventral hippocampus	EMG electromyogram neck
YEUX eye movements	PLETH volume changes front leg

and Jasper, 1937) Their onset is accompanied by neurological signs that the subject is in a deeper condition of sleep, contrary to what the EEG asynchronization suggests There is complete relaxation of antigravity muscles, disappearance of stretch reflexes, rapid movements of the eyes (abbreviated as REM for "rapid eye movements"), together with movements that may include twitching, jerking, or even mild convulsions (Fig 9 117) In laboratory or domestic animals one may see moving of the vibrissae, extremities and tail Everyone who has observed the sleeping state of a pet cat or dog has noticed this type of movement The loss of tone in antigravity muscles and the absence of stretch reflexes serve to distinguish this condition from any tendency towards wakening Furthermore, the threshold for arousal is higher than when the EEG shows the slower type of waves This condition, known as "fast sleep"—sleep with fast waves—is also called *paradoxical sleep* because of the inconsistency between the neurological condition and the EEG wave forms It is one of the deepest stages of sleep (Jouvet, 1967)

SLEEP WITH SLOW WAVES

In most of the experimental literature devoted to the subject of sleep, the condition studied was the one characterized by large, slow waves in the EEG, the apparently synchronized type of electrical activity It has been analyzed in experimental animals after surgical transection of the neuroaxis, as well as with the aid of the stereotaxic instrument for placement of lesions, for stimulation or for recording Modern work on this problem began just before World War II with a publication by the Belgian neurologist, Frederic Bremer (1936) He transected the brainstem of cats in the midcollicular region of the mesencephalon, thereby creating what he referred to as a *cerveau isole* preparation (with "isolated mind") The cerebral cortex of such an animal appears to be continually asleep as judged by the EEG pattern, and it cannot be activated by at-

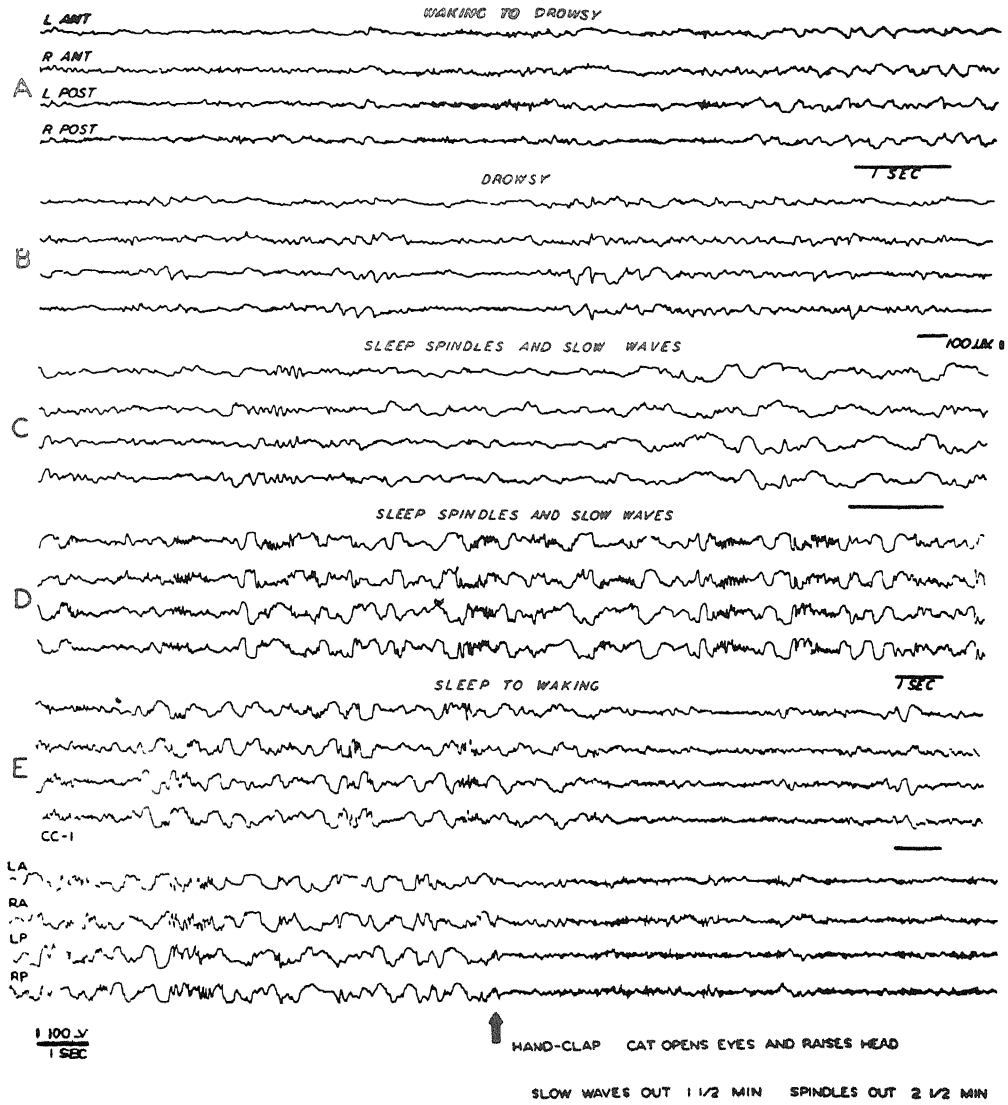


Fig 9 118 *A–E.* Electrocorticogram of cat as it passed from waking to sleep and then awakened again *Bottom* Cat awakened by hand-clap (Magoun 1952)

tempts to arouse the animal On the other hand, if the plane of transection is at the lower border of the medulla oblongata (*encephale isolé*, "isolated brain"), the EEG may show the waking pattern, alternating with periods in which the pattern is characteristic of sleep Moreover, the sleep pattern can be converted to the waking pattern by sensory stimulation via cranial nerves. (One should note that the preparation with *encephale isolé* requires artificial respiration because the medullary respiratory centers are cut off from respiratory motor mechanisms in cervical and thoracic cord)

Bremer's interpretation of his findings was that the sleeping state of the *cerveau isolé* preparation was the result of removal of sensory input as a result of transection of ascending pathways between brainstem and cortex In the preparation with the lower level of section, by contrast, the ascending pathways from sensory portions of cranial nerves remain intact and capable of stimulating higher levels including thalamus and cortex Later experi-

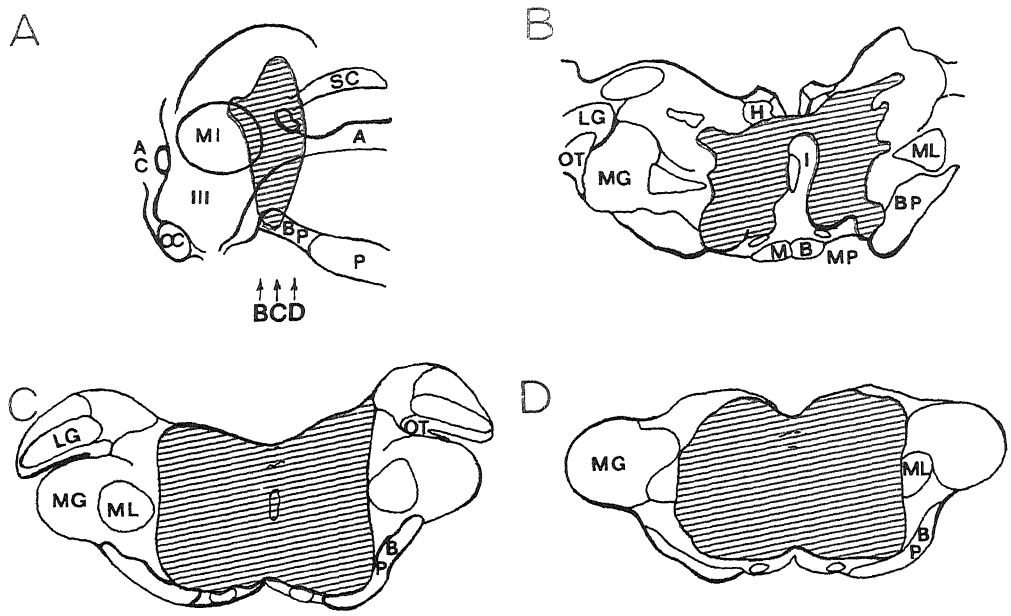


Fig 9 119 Midsagittal (*A*) and cross sections (*B–D*) through brainstem of cat with destruction of central portions of reticular formation (Lindsley et al 1950)

III third ventricular
AC anterior commissure
L lat geniculate body
MI massa intermedia
OC optic chiasm
P pons
A aqueduct
BP basis pedunculi
MG medial geniculate body
ML medical lemniscus
OT optic tract
SC superior colliculus

ments showed, however, that mechanisms other than the sensory systems are necessary for preservation of the waking state

Reticular Activating Mechanism. This latter, alternate interpretation came from experiments in which Lindsley and his associates (1950) performed stereotaxic operations that only partially transected the brainstem Certain of the lesions were placed laterally so that they interrupted the specific or lemniscal sensory pathways on each side Others were placed medially in the central portion of the reticular formation on both sides of the midline (Fig 9 119) In animals with the lateral lesions a normal transition from sleeping to waking state occurred as judged both from the behavior of the animal and from the EEG But in animals with medial lesions the sleeping state was prolonged as in Bremer's *cerveau isolé* preparation, with one significant difference Because the main sensory pathways were intact following the medial lesions, sensory stimulation would activate the cortical electrical pattern momentarily Then when the stimulation was ended the cortex lapsed into a sleeping pattern (Fig 9 120) These observations confirmed the ability of sensory input to alert the cortex as Bremer had inferred, while they showed also that the maintenance of an alert state between stimulations depends upon projections from the medial reticular formation to higher levels of the brain

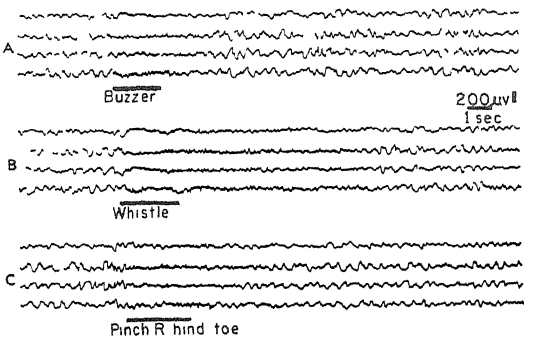


Fig 9 120 EEG of cat with lesion shown in Figure 9 119 Record shows persistent sleep pattern with brief transition to waking pattern during and just after stimulation by sound or pain Four different cortical regions show similar wave forms (Lindsley et al , 1950)

Further data on this question have been obtained by the technique of stimulation Mo-

ruzzi and Magoun (1949) found that electrical stimulation of the medial reticular formation transforms a high voltage, slow wave EEG into the low voltage, fast wave or activated pattern Because it has this property of alerting the cerebral cortex this function of the reticular formation is ascribed to the *ascending activating system* Its influence is transmitted via multisynaptic pathways via the nonspecific thalamic nuclei, and thence to the cerebral cortex (Chap 2, p 9-40) The portions of the reticular formation that are concerned with activation of the cortical EEG are, to some degree, under the control of the cortex Segundo (1956) and his colleagues have shown that electrical stimulation of certain cortical regions will produce an EEG desynchronization which appears to be identical with the desynchronization observed by Moruzzi and Magoun (1949) after electrical stimulation of the reticular formation In a feedback loop the ascending activating system is stimulating cortical alertness, but the cortex is also projecting back to brainstem to modify the output of the activating mechanism The presence of this feedback pathway was demonstrated by Jasper et al (1952), and by French et al (1955), who recorded evoked potentials from the reticular formation following stimulation of a variety of areas in the cerebral cortex

The fact that the lateral hypothalamus is anatomically and functionally continuous with the reticular formation was first suggested by studies on sleep A number of conditions that occur clinically are typified by somnolence, prolonged sleep, or even coma (chap 9, p 9-177), in patients with lesions in the hypothalamus or mesencephalon The lesions may be caused by an infection as in "sleeping sickness" (von Economo, 1929), or may be a result of pressure from a tumor, or of hemorrhage or occlusion of a blood vessel Ranson (1936 to 1937) and his associates found that injury to the caudal part of the lateral hypothalamus produces in monkeys a somnolence that appears to be identical with the clinical condition. Stimulation of this region through chronically implanted electrodes causes animals to become alert and active

The current hypothesis, therefore, is that the waking state is initiated and maintained by projections from the reticular formation (including lateral hypothalamus) via nonspecific pathways through thalamus to cerebral cortex (Fig. 9 121) The next question is, what turns

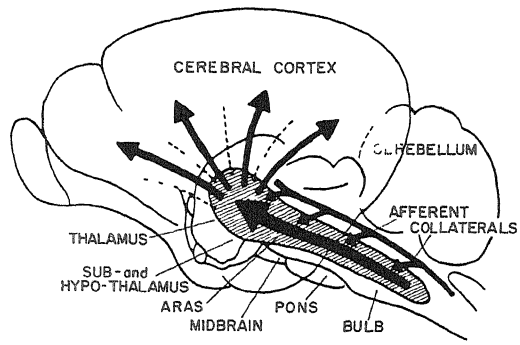


Fig 9 121 Outline of the brain of the cat showing general location of the activating mechanism in the tegmentum of the brainstem and its relation to principal afferent systems and projections via thalamus to cerebral cortex Starzl et al cited from Magoun 1952

on the reticular activating system? This question, too, has been investigated by authors who have stimulated the several parts of the brainstem Pioneer work was done by the Swiss neurophysiologist, W R Hess, who invented a technique for implanting electrodes in anesthetized cats in a manner that permitted stimulation later after the animal had recovered from the operation (Fig 9 6, p 9-9) In the course of experiments begun in 1925 and summarized later (Hess, 1954), he discovered that the application of a particular type of stimulation to the thalamus lateral to the ventral portion of the massa intermedia would produce sleep From these results he concluded that the brain contains a mechanism that induces sleep in an active sense A series of papers by Akert et al (1952, see also R Hess, Jr, et al, 1953) has essentially confirmed the earlier observations, and extended them by showing the expected correlations between the state of the animal and the EEG Their data imply that the sleep-waking mechanism, like other components of the central nervous system as described in preceding pages, is a dual system It may have a portion that turns on the waking state, and another that "turns off" waking and so induces sleep If so, it fits into the pattern of reciprocal interaction of alternating inhibition and facilitation that can be identified at all levels of the neuroaxis

Hypothalamic Suppressor Mechanism Evidence as to the existence and location of this second component of the control system has come also from experiments in which the opposite of somnolence has been observed after operations upon the base of the brain

Although Hetherington and Ranson (1942) did not report any observations upon sleep as such, they did observe a condition of extreme, prolonged, and "frantic" hyperactivity in rats just after operations upon the tuberal hypothalamus Later these rats became almost completely inactive With lesions placed more rostrally in the suprachiasmatic or preoptic region, however, Nauta (1946) observed both hyperactivity and lack of sleep McGinty and Sterman (1968) have recently described similar changes in cats with large, bilateral suprachiasmatic lesions, and have correlated the behavior of the animals with their EEG patterns The largest lesions abolished sleep, smaller ones significantly decreased the duration of both the slow-wave and paradoxical stages of sleep Taken together these data suggest that, if there is a dual mechanism for the sleeping-waking control, one part lies ahead of the other The caudal hypothalamus and reticular formation seem to contain the activating component, whereas the rostral hypothalamus and preoptic region include the suppressor The relation of these results to those of Hess and others as presented in the preceding paragraph, however, is not yet clear

PARADOXICAL SLEEP

As noted above, paradoxical sleep or REM sleep is characterized by low voltage fast EEG, profound atonia of the somatic musculature, and rapid conjugate eye movements In addition, recordings from deep structures and cortex show a θ (4 to 7 cycles/sec) pattern in the hippocampus and intermittent bursts of activity in the pons, the lateral geniculate body, and visual cortex Jouvet (1967) has classified the phenomena of paradoxical sleep into two categories, tonic and phasic Tonic phenomena include the EEG, hippocampal θ rhythms, and muscular atonia Phasic phenomena include the rapid eye movements, the bursts of activity in the pons-geniculate-cortical complex, and occasional muscle twitches In a series of studies in which lesions were placed in a variety of regions, Jouvet has determined that the structures apparently responsible for the phasic components of paradoxical sleep are the rostral part of the nucleus reticularis pontis candalis and the caudal part of the nucleus reticularis pontis oralis Stimulation of these regions in the tegmentum of the pons during normal sleep produces paradoxical sleep

More recently Jouvet has also found that bilateral lesions in the region of the locus caeruleus will suppress or abolish the tonic component of paradoxical sleep without affecting the phasic component Animals in which the tonic component has been abolished by such lesions show considerably increased muscle twitching during phasic paradoxical sleep

HUMORAL MECHANISMS OF SLEEP AND WAKEFULNESS

Legendre and Pieron (1910) suggested that an unknown substance, which they called hypnotoxin, found in the cerebrospinal fluid (CSF) is involved in producing sleep They observed that CSF removed from fatigued animals and injected intracisternally into normal, rested animals tended to induce sleep Schnedorf and Ivy (1939) replicated these experiments with additional controls Dogs that received CSF from fatigued dogs apparently went into a deep sleep Control dogs, which received CSF from rested dogs, or injections of their own CSF, however, also showed some depression but to a lesser extent Many other authors have attempted similar experiments Their results have been summarized by Jouvet (1967) in the suggestion that serotonin is the neural transmitter that induces sleep

The existence of an activating substance was suggested by Purpura (1956) He connected each femoral artery of a cat to a femoral vein of a second cat and vice versa to obtain intermixing of the total blood volume Stimulation of the medial bulbar reticular formation produced immediate cortical activation in the stimulated cat After about 60 sec an activation pattern could be seen in the electrocorticogram of the recipient cat Unfortunately, Purpura was unable, on the basis of available evidence, to advance an hypothesis concerning the identity of the circulating activator substance

More recently, Monnier and Hosli (1965) have reported that dialysate from venous blood of sleeping or active donor rabbits will produce sleep or alert behavior, respectively, when injected into the circulation of recipient rabbits

The more recent studies of paradoxical sleep have suggested humoral mechanisms controlling this form of sleep also Among the more ingenious studies in this area was one by Jouvet (1967) and his colleagues who devised a method of differentially depriving animals of

paradoxical sleep but not slow sleep Jouvet placed cats on a brick in a shallow pool of water for periods of days The animals maintained sufficient muscle tone during slow sleep to be able to keep from falling into the water During paradoxical sleep, muscle tone decreased to the point that the animals fell into the water, waking up Animals deprived in this fashion of paradoxical sleep seemed to build up a need for paradoxical sleep, for when allowed to sleep normally the percentage of sleeping time devoted to paradoxical sleep was far greater than normal This build-up of a need for paradoxical sleep is attributed to the accumulation of an unidentified substance during paradoxical sleep deprivation Related studies in which humans were deprived of paradoxical sleep, by being awakened whenever EEG and eye movement records indicated paradoxical sleep, also show the same build-up of need for paradoxical sleep This accumulation can be suppressed by monoamine oxidase inhibitors For example, administration of Nialamide to animals deprived of paradoxical sleep will prevent the subsequent appearance of paradoxical sleep 5-Hydroxytryptophan will also suppress paradoxical sleep γ-Butyrolactone in doses of approximately 50 mg /kg will produce paradoxical sleep and large doses of Reserpine will differentially produce the phasic phenomena of paradoxical sleep On the basis of such evidence, Jouvet (1967) has suggested that the substance accumulating during paradoxical sleep deprivation may be an amine, probably norepinephrine

SIGNIFICANCE OF SLEEP

Every thoughtful person has speculated as to the function of sleep Yet almost nothing is known of it scientifically It is clear that in the absence of sleep the more subtle functions of the nervous system are affected first, with changes in mood, motivation, and attention (Wilkinson, 1969) With more prolonged sleep deprivation even the subcortical, more nearly automatic functions of the brain suffer, and symptoms such as tremors, nystagmus, ptosis, and dysarthria occur (Sassin, 1970) As Hobson (1969) has pointed out, there are certain forms of illness that are most likely to occur when the patient is asleep, they include nocturnal angina, asthma, and emphysematous anoxia

One may suppose that neurons are not capable of continuous, repetitive discharge, although what happens to limit this capacity is unknown It might be some limitation of energy transduction, or it might be the accumulation of ions during activity that must be extruded during recovery, or it might be some as yet immeasurable change in the properties of one or more of the neuron's membrane systems Those cells that are active in both waking and sleeping states, such as the respiratory system, are not continuously active, rather, they have their own cycles of activity and rest, such as the cycle that includes inspiration-expiration-inspiration, etc Most of the neural control systems exhibit some form of rhythmicity like this, or at least some alteration in the level of their discharge Not so the activating mechanism Unless it has some mechanism for operating on a basis of "taking turns," it seems to be discharging more or less continuously so long as the individual is awake It appears that neurons are not constituted so as to permit this kind of continuous activity indefinitely

Emotion

How we "feel" is so important to everyone, and disorders of emotional state are so prevalent and troublesome, that any possible understanding is desirable both for the insight it may provide into human behavior, and for its practical value in clinical medicine In the same manner in which one can distinguish between movement and a desire to move, it is possible to separate emotional expression from emotional experience Emotional expression, like a movement of the body can be studied in laboratory animals, in infants incapable of speech, or in other circumstances where verbal or written communication is not possible To attempt to analyze volition or the experience of an emotion, by contrast, requires either that the observer himself be the subject of the investigation, or that the subject be able to report to the observer just what he believes to have taken place There are, therefore, two almost completely separate kinds of literature on the neurophysiology of emotion One is concerned with animal behavior following operations upon the nervous system, the neurohistology in such reports tends to be precise and even elegant, but the nature of the emotional experience is uncertain because it can be inferred only from an anthromorphic point of view On the other hand, reports based upon the experience of human patients with lesions of the central nervous system may be highly

informative as to the kinds of emotion the individuals have undergone, but often are deficient in precise neurological localization, because such lesions usually are not localized sharply enough to permit the assignment of any one function to any specific structure

EXPERIENCE VS EXPRESSION

There is no certain knowledge as to mechanisms that create *awareness* of the state of emotional tone This problem is akin to the question of how does the individual finally become aware of any type of sensory or motor experience Perhaps the cerebral cortex contains a scanning mechanism that at some frequency surveys and registers the overall patterns of activity in all regions If so, then what scans the scanner?

The *expression* of emotional states, on the other hand, can be attributed fairly reliably to mechanisms that already have been mentioned, including *a*) those cortical regions from which autonomic responses can be evoked upon stimulation, *b*) the intermediate or transitional portion of the neuroaxis known as the limbic system (see later), and *c*) the hypothalamus and its projections to brainstem and spinal cord Bard (1928) suggested that these portions of the brain, but especially the hypothalamus, are responsible for "integration" of the somatic and visceral phenomena that characterize behavior that accompanies emotion The functions of these neural mechanisms have been studied in different species of laboratory animals, with reference to several typical emotional states (see Pribram, 1969, Black, 1970) The one that is perhaps the best understood is *rage*, mainly because early investigators found that it can be evoked easily under laboratory conditions, as reviewed in the following paragraphs Other motions that have been studied include *placidity*, what appears to be *pleasure* (as judged by purring of cats, and other similar behavior), and emotions presumably associated with behavior patterns of *reproduction* The list is scarcely a satisfactory one, when it is contrasted with the range and variety of human emotional tone

SHAM RAGE

It was first demonstrated by Goltz and Ewald, (1896) that the reactions which usually accompany displeasure and anger are more readily evoked in an animal deprived of its cerebral cortex (decorticated) than in the normal animal In Goltz's classical experiment the hemispheres and a large part of the thalamus were removed from a dog The disposition of the animal was greatly altered by the operation, it being very readily aroused to anger Barking, growling, baring the teeth, or snapping occurred upon the least provocation Cannon and Britton (1925) produced a similar state in cats by removal of the cortex, the decortication being performed by means of a pointed stilet inserted through the orbital cavities Immediately following recovery from the anesthetic the animals showed the following remarkable phenomena which these observers termed *sham rage*—lashing of the tail, erection of the hairs, protrusion of the claws, dilatation of the pupils, sweating, struggling and biting, greatly increased rate of respiration, and a rise in blood pressure Most of these manifestations will be recognized as being of sympathetic origin In many instances the picture was a combination of fear and anger Liberation of epinephrine, as indicated by an increase in the rate of the denervated heart, also occurred The mildest stimulus such as jarring the table or lightly touching the animal was sufficient to evoke a paroxysm of rage The condition was designated as "sham" or as "quasi," because these investigators believed that true rage cannot be experienced unless higher levels of the cerebral cortical mechanisms are involved

In decortication experiments Bard (1928) found that the posterior and ventral portions of the diencephalon are responsible for the development of sham rage Typical behavior occurred after decortication and section through the hypothalamus at about the middle of the tuber cinereum Rage also was noted after an operation that removed the basal ganglia and the dorsal half of the diencephalon, i e , the thalamus, but left the hypothalamus connected with the midbrain The condition failed to appear if the section separated the caudal part of the hypothalamus from the brainstem as a high level decerebration

HYPOTHALAMIC RESTRAINING MECHANISM

Another variety of rage was discovered by Wheatley (1944) when he was studying hypothalamic hyperphagia in cats He noted that bilateral lesions in the ventromedial region of

the hypothalamus not only lead to overeating and obesity, but also to a behavior pattern in which the animals are " extremely, chronically and incurably savage in behavior and expression" (Ingram, 1952, p 401) These animals are so vicious that it is all but impossible for laboratory personnel to handle them Apparently they cannot be tamed once the lesions have been made Moreover, the change in behavior from the friendliness of a pussycat to this malevolent wildness seems to occur instantaneously because it is evident immediately after recovery from the anesthesia of the operation Although it is a hypothalamic injury that leads to this syndrome, this is not sham rage and seems to have little in common with the quasi-emotional state observed by Goltz, by Cannon, and by Bard and their associates There is nothing that can be called pseudo, sham, or quasi about the performance of a cat with ventromedial lesions The animal seems to possess full cortical function, both sensory and motor, and gives every evidence of experiencing the emotion it expresses Its rage is perfectly directed at whomever disturbs it, whereas sham rage cannot be directed and is therefore relatively innocuous Compared to the extensive ablations that lead to sham rage, the lesions that are effective in the ventromedial region are tiny Thus far, the syndrome is known to appear only in cats, rats with similar lesions do not show similar behavior to the same degree, and perhaps not at all, whereas macaque monkeys with ventromedial lesions are reported to be obviously more tame and friendly the morning following the operation (Hamilton, C L, personal communication)

CORTICAL INHIBITORY MECHANISMS

Both of these types of rage seem to be analogous to decerebrate rigidity in that they are caused by over activity of mechanisms that the surgical operation has released from an inhibition or restraint usually provided by the regions that were taken away In Wheatley's experiments the ventromedial nuclei of the hypothalamus seemed to be the origin of the restraint, or perhaps a vital part of the pathway by which it is exercised when the brain is intact. In the case of the sham rage, however, the inhibition seems to come from certain critical regions of the cerebral cortex. Thus, Bard and Mountcastle (1948) have shown in cats that the restraining influence is exerted in certain parts of the rhinencephalon (the cortex of the pyriform lobe, amygdaloid nucleus and hippocampal formation) and an area of cortex lying ventral to the cingular sulcus in front of and below the rostrum of the corpus callosum The cortex of this cingulate gyrus and the hippocampus are portions of the limbic system, a transitional zone between the rhinencephalon and the neocortex Following operations that removed the neocortex alone, leaving the rhinencephalon and transitional cortex uninjured, the cats showed a completely different kind of behavior They were extraordinarily placid, responses that seemed to express pleasure were predominant and often exaggerated Procedures that in a normal animal provoke struggling, clawing or biting, or an attempt to escape, in these cats evoked no sign of anger or even of displeasure The animals responded to mild nociceptive stimulation by purring. Subsequent removal of those parts of the limbic system mentioned above transformed these animals from a state of placidity into one where upon the slightest provocation they exhibited all the signs of rage They were in this respect, therefore, similar to the cats of Wheatley's experiments

It is apparent from these experiments that the cortex and other regions of the brain exert a restraining influence upon those mechanisms of caudal hypothalamus and brainstem that govern the expression of displeasure, resentment, and anger, at least in cats and dogs

Similar operations performed upon monkeys have given results that are in some respects different, and that are even more interesting than the experiments in which cats and dogs were used Kluver and Bucy (1939) reported that bilateral extirpation of the temporal lobes in monkeys, including both the neocortex and the rhinencephalon on the medial surface of the lobes, gave rise to a group of symptoms that has become known as the Kluver-Bucy syndrome It includes the following *1) Visual agnosia.* These animals seem to have lost the ability to recognize and to detect the meaning of objects on the basis of visual criteria alone There also seems to be a similar agnosia in the auditory and tactile fields *2) Oral tendencies* There is a very strong tendency to examine all objects with the mouth *3) "Hypermetamorphosis."* There is a marked tendency to take notice of and to attend to every visual stimulus *4) Tameness* These animals seem to

have lost their sense of fear, and there is a marked diminution in their emotional responses 5) *Hypersexuality* There is a striking increase in various sexual activities and in their diversity This change does not occur until several weeks after the operation Females may show a complete lack of maternal behavior 6) *Changes in dietary habits* Monkeys do not ordinarily eat meat but after the removal of both temporal lobes they do so without hesitation There is also an increase in the amount of food consumed In a few cases some of these symptoms have been observed also in human beings following bilateral temporal lobectomy

In later experiments by other authors it was possible to attribute some of these changes to specific structures within the temporal lobe For example, removal of the amygdala and overlying pyriform cortex is said to produce placidity (Schreiner and Kling, 1953) The hypersexuality has been attributed to lesions of the pyriform cortex without injury to the amygdala (Green et al , 1957) And it has been suggested that the psychic blindness may be the result of removal of the temporal neocortex, or perhaps of connections between the visual system and the posterior regions of the limbic system

LIMBIC SYSTEM

Fibers in the olfactory tracts terminate in the medial and lateral olfactory striae of the mesial and orbital cortex, from which there arise projections that are so variegated and that extend to so many other regions of the brain that the whole complex has been classified as a distinct portion of the nervous system, the *rhinencephalon* or "nose brain " In 1933 the American neuroanatomist, C Judson Herrick, surveyed the possible functions of the parts of the cerebral cortex belonging to this olfactory brain, and noted that they are "bound up with" other neural systems including those for feeding, nutrition, and reproduction Furthermore, the olfactory sense gives warning of danger, and guides an animal to a mate, to food, and "other desirables " He also observed that these portions of the brain are remarkably constant in their organization from lowest to highest of the vertebrates From these and other considerations he suggested that the olfactory system has two functions, *a*) adjustments in other systems so as to secure external orientation of the body, and *b*) an action " upon the internal apparatus of the general bodily attitude, disposition, and affective tone " (p 14)

This latter inference of Herrick was made more specific in 1937 by Papez in the form of a proposal that the principal functions of most of the supposed components of the rhinencephalon are not directly concerned with olfaction, but rather with emotional experience To reach this conclusion Papez drew upon the two types of evidence mentioned above He reviewed what was known of the location of lesions that in human patients were associated with emotional disorders, and he summarized also the results of Cannon, Bard, Ranson, and other authors (see above) whose work had disclosed the role of the hypothalamus in the expression of emotion such as rage, and in other autonomic patterns From these two lines of research Papez decided that the anterior nuclei of the thalamus, the cingulate gyrus, the hippocampus, and the interconnections of these regions constitute " a harmonious mechanism which may elaborate the functions of central emotion" He further suggested that the expression of the emotion is brought about via projections from these structures to the hypothalamic mechanisms other authors had described (Fig 9 122) In 1952 these portions of the brain were given the name, *limbic system* by MacLean (1967), because Broca had used that word to designate the portion of the

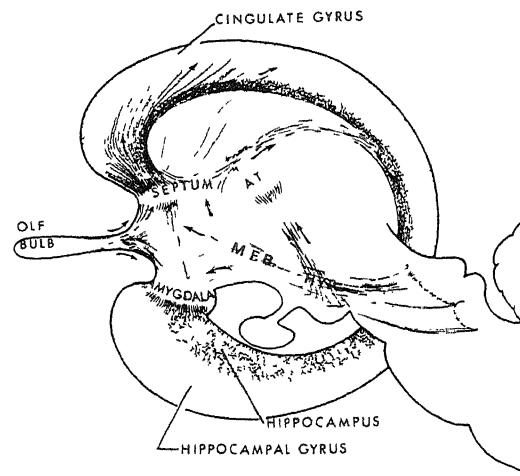


Fig 9 122 Principal interconnections of rhinencephalon with diencephalon and limbic cortex *AT* anterior nucleus of thalamus, *HYP* hypothalamus *MFB* medial forebrain bundle of fibers, they run through the lateral hypothalamus Cited from MacLean 1967

cortex that has the character of a "border" as it surrounds the brainstem The limbic cortex is intermediate in position between the portions of the brain that are common to even the most primitive vertebrates, and the neocortex that characterizes primates and man MacLean pointed out that in the cortex the levels of function cannot be represented by horizontal or coronal planes of section, but rather are arranged concentrically, with the most primitive in and adjacent to the brainstem, and the most advanced in the most superficial location (Fig 9 123) In the limbic system the principal cortical structures are the cingulate and hippocampal gyri, and the hippocampus (Fig 9 124) Subcortical portions include the amygdala (also classified as one of the basal ganglia) and the region known as the septum This latter is found on each side near the midline, a part of the telencephalon that is virtually continuous with the dorsal and rostral hypothalamus Stimulation of the cortex belonging to this system in anesthetized animals evokes autonomic responses such as changes in blood pressure, heart rate, gastrointestinal motility and secretion, as well as in pulmonary ventilation (Kaada et al , 1949) In animals not anesthetized this stimulation results in " pleasure, grooming and sexual reactions seen in courtship behavior" (MacLean, 1970, p 134) The connections of these structures with other parts of the brain have been followed histologically, and also by means of the technique of evoked potentials

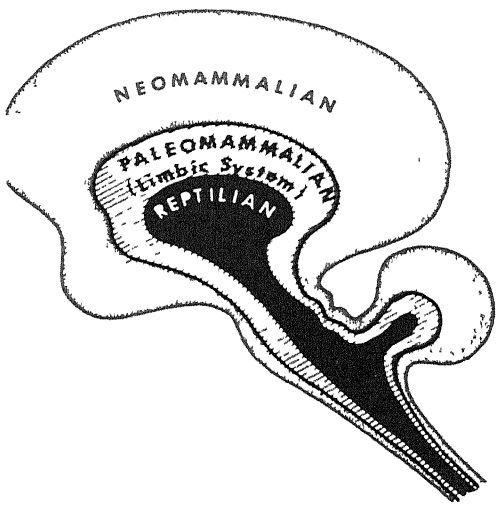


Fig 9 123 Hypothetical plan of organization of three levels of cortical integration named with reference to comparative neurophysiological criteria (MacLean 1967)

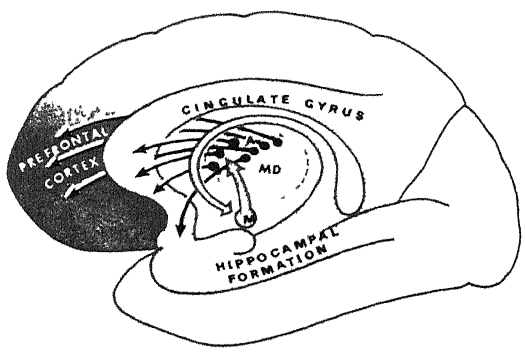


Fig 9 124 Relationships of thalamus with cerebral cortex, including cingulate gyrus and prefrontal cortex *A*, anterior nucleus of thalamus, *MD*, dorsomedial nucleus of thalamus, *M* mammillary body (MacLean, 1967)

Human patients with lesions in these areas are said to undergo a form of epilepsy (psychomotor) where the symptoms are autonomic discharge and emotional experience, stimulation during neurosurgical operations by way of electrodes applied to these structures evokes responses comparable to the psychomotor epilepsy (Penfield and Jasper, 1954) In a few persons recordings have been made, and stimulation has been applied, through implanted electrodes, while they are living a more or less conventional life as hospital patients The data so obtained are in general agreement with results obtained from experimental animals For example, "spontaneous" bursts of anger were evoked by stimulation of the right amygdala Other kinds of experience elicited by the stimulation included pleasant sensations, elation, deep thoughtful concentration, "odd feelings," relaxation, and colored visions (Delgado, 1970) Delgado concluded from these observations that stimulation of this type " may have further therapeutic use in order to obtain beneficial modifications of emotional reactivity" (p 201) He also stated that it raises ethical and philosophical questions

Speech and Its Disorders

Communication is a word used in physiology with two slightly different meanings, the first of which includes the second In the more generalized sense it denotes the interaction of any two units in a fashion that modifies the function of either It may be applied to the influ-

ence of one neuron upon another, or to the stimulus transmitted between cells by a hormone, as well as to communication between persons What is transmitted is *information* In the more restricted sense communication requires signals that are created by one individual and perceived by another via sense organs The signals may be physical for optical, aural, or thermal detection, or they may be chemical The information they transmit may be extremely complex, even in relatively simple animals Perhaps the best studied example in lower forms is the ability of bees to inform other inhabitants of a hive as to the direction and distance at which a specific food source may be found (von Firsch, 1967) Ethologists and psychologists have studied many other similar kinds of communication among other animals (Dethier and Stellar, 1970)

Speech and its relatives, writing and reading, are the most elaborate forms of communication, and are limited to the human species Cortical function is necessary for speech, according to the observations of Penfield and Roberts (1966) it is the cortex of the left cerebral hemisphere that is usually responsible for speech, even in left-handed persons where the right hemisphere is said to be "dominant" for motions such as writing and other skilled performance (Fig 9 125) Geschwind (1970) has stated that, "Contrary to generally accepted views, there is a striking anatomical difference between the temporal speech region on the left side and the corresponding region of the right hemisphere "

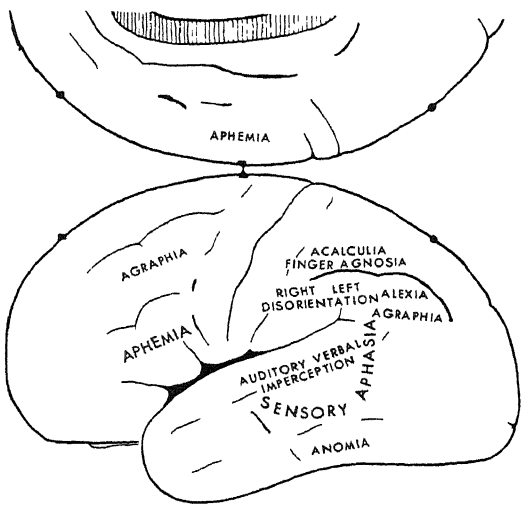


Fig 9 125 Summary of the location of cortical lesions which various authors have associated with disorders of speech (Penfield and Roberts 1966)

It is supposed that the first stage in the development of speech is the association of certain sounds—(words)—with visual, tactile, and other sensations aroused by objects in the external world After definite meanings have been attached to certain words, pathways between the auditory area of the cortex and the motor area for the muscles of articulation become established, and the child attempts to formulate and pronounce the words which he has heard This act of verbal expression involves the coordinated movements of a large group of respiratory, laryngeal, lingual, pharyngeal, and labial muscles Later, as the child learns to read, auditory speech is associated with the visual symbols of speech, and finally, through an association between these and the motor area for the hand, the child learns to express his auditory and visual impressions by the written word (see Lenneberg, 1970)

APHASIA

Aphasia is the term applied to disorders of expression in speech, writing or signs and symbols, as well as to disabilities in comprehension of spoken or written language Lesions causing aphasia are usually in the cerebral cortex, but they do not necessarily involve the motor mechanisms of area 4, the corticobulbar fibers, cranial nerve nuclei nor peripheral nerves Rather, the lesions typically lie in cortical "association" areas (Fig 9 125) near what are presumed to be the highest levels of cortical function Nevertheless, in spite of this apparent juxtaposition of localization, the intelligence of a patient may be unimpaired in the presence of severe aphasia This implies that there are functions of the cortex at levels higher than the speech mechanisms

The faculty of speech is based upon highly complex neural systems They can be studied in two ways, by analysis of defects in patients with naturally occurring lesions of the cortex or following operations, or by stimulation of the cortex at an operation in a patient given only local anesthesia for the surgical procedure The fact that this electrical stimulation is categorically different from natural neural activity is dramatically shown by the observation that in an area where a lesion causes aphasia, electrical stimulation when a patient is speaking will arrest the speech After the

stimulus is ended the patient may report that he does not know why he stopped talking, but he was unable to think of the words he wished to use There are four such areas in the left hemisphere, the dominant hemisphere in a right-handed person They include the following *1) lower frontal* cortex of Brodmann's area 4 (known as Broca's area), *2) upper frontal* on the mesial surface of the hemisphere anterior to the foot area, *3) parietal*, posterior to the lower part of the postcentral gyrus, and *4) temporal*, in the posterior part of the temporal lobe A lasting injury to any one of these areas causes persistent aphasia, with the possible exception of the one in the upper frontal region

Responses that seem to be related to speech can be obtained following electrical stimulation of other cortical regions For example, in a conscious patient a stimulus applied to the lower part of the precentral gyrus in the lips, jaw, and tongue areas of either hemisphere causes the emission of a crude sound like the cry of an infant or of an epileptic at the beginning of an attack This sound does not resemble a spoken word Similarly, stimulation in the supplementary motor area on the mesial surface of either hemisphere causes vocalization, perhaps less crude than the former but definitely not speech These kinds of stimuli are believed to be transmitted to the cortex of area 4 in a pattern that evokes vocalization as a motor act independent of the higher cortical associations necessary for speech Ablation of either of these areas causes only a temporary speech defect, perhaps motor in nature One should add, however, that operations upon the lower part of the precentral gyrus is the dominant hemisphere may cause temporarily a true aphasia This is believed to represent some interference with the function of area 44, possibly as a result of interruption of its blood supply

Classification of the Aphasias. There is no one system for classifying aphasias, the several neurologists who have studied this phenomenon having usually proposed categories that seem consistent with their own findings In this account the views of Sir Henry Head (1926), an English neurologist, are followed (Fig 9 126) Head's studies, done upon patients suffering from gunshot wounds of the cerebral cortex, led him to the conclusion that most aphasias could not be identified as

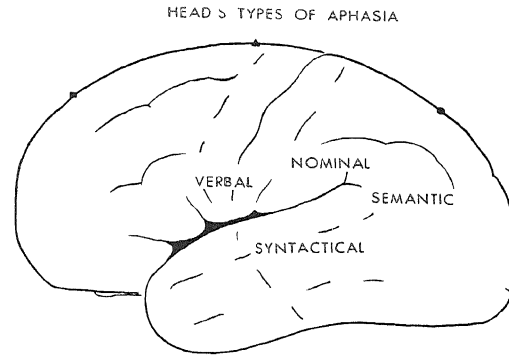


Fig 9 126 Brain map to show locations of cortical lesions responsible for types of aphasia described by Head (1926) From Penfield and Roberts 1966

simply motor or sensory in type because each patient exhibited multiple or combined deficiencies He concluded that speech mechanisms are more diffusely represented than many authors have supposed, and that the several cortical speech "centers" do not have discrete and independent functions He noted that the more complicated or abstract the idea which must be understood or expressed, the greater is the difficulty Thus, an aphasic patient may be able to name objects correctly but fails to find the word for a more abstract idea, e g , color Shown a black object, for instance, and asked to name its color he fails to do so, yet indicates that he recognizes that it is black by saying "what you do for the dead " An aphasic soldier when shown a red object said, "what the staff wear "

Head divided aphasia into four types, as follows

1) Verbal Defects The outstanding feature is that the power to express an idea in words is practically lost The patient, however, is not entirely speechless but can usually utter a few monosyllables, "yes" or "no," etc , or ejaculations and emotional expressions, such as, "damn," "oh dear me " When the disorder is less severe, the words are mispronounced but sentences are correctly constructed For such patients reading is difficult, and writing is defective or impossible They usually understand printed or oral commands

2) Syntactical Defects (Agrammatism, Jargon Dysphasia) The patient is voluble but speaks a jargon in which, though the individual words may be fairly accurately pronounced, they are strung into short phrases or badly constructed sentences without articles, prepositions, or conjunctions The ability to read

aloud is impaired, and the understanding or ordinary conversation is defective

3) Nominal (Naming) Defects In this form of speech disorder the patient has difficulty in finding the right word to express his meaning or for naming a well-known object He will often employ a descriptive phrase in substitution for the word which he cannot recall These patients can draw from a model either directly or from memory, after it has been shown and then removed, but are usually unable to draw from imagination They write a coherent letter with difficulty, usually fail to carry out simple arithmetical exercises, and confuse the values of coins

4) Semantic Defects A patient suffering from this type experiences little difficulty in articulate speech, can name objects, understands invividual words and some sentences, but the general meaning of what he hears escapes him He often fails to follow his own utterances to an intelligent conclusion, his sentences trailing off as though he had forgotten what he had started out to say When shown a picture he picks out the details but fails to grasp the meaning which it conveys to others Such a patient therefore misses the point of a joke whether this is printed, told to him, or is in pictorial form He fails to comprehend the significance of much that he sees and hears There is no impairment in the pronunciation of words and, though speech tends to be in short jerky sentences, syntax and intonation are not disturbed

Head discarded the conception that the neural basis of speech consists of strictly localized anatomical "centers" wherein resides exclusively one or other of the speech functions —auditory, visual or motor He believed, however that regions exist in the cortex "where the progress of some mode of action can be reinforced, deviated or inhibited " These regions constitute foci of integration—convergence points for association paths Destruction of one or other of such foci or "knots" of association paths will depress *as a whole* the psychological processes underlying speech Nevertheless, the nature of the speech disability is influenced by the particular region of the cortex involved If the lesion is in the neighborhood of the lower part of the precentral and postcentral convolutions of the dominant hemisphere the speech defect tends to be of the *verbal* type In injury to the temporal lobe the speech defect tends to be of the *syntactical* type In a lesion in the region of the angular gyrus of the dominant hemisphere the patient has difficulty, particularly, in finding names for things (*nominal defect*), damage to the cortex in the region of the supramarginal gyrus results in a *semantic defect*

OTHER DISABILITIES OF SPEECH AND PERCEPTION

Anarthria or dysarthria is loss or difficulty of speech due to paresis, paralysis or ataxia of the muscles concerned in articulation There is no impairment of the psychical aspects of speech, i e , "internal speech" is unaffected, there is no difficulty in the comprehension of spoken or written speech Other functions, e g , swallowing, which are dependent upon the same groups of muscles as those used in speech, are also frequently affected

Apraxia (unable to act) is the inability to perform purposeful movements at will, either at command or in imitation, though the muscles normally engaged in the act are not paralysed It is allied to aphasia, which might be called apraxia of the speech faculty Apraxia may be sensory or motor In the former, the patient does not recognize the significance of an object and therefore cannot put it to its proper use, this is *visual agnosia* In motor apraxia the patient has no conception, or a very defective one, of the pattern of muscular movement required to perform a purposeful act For example, apraxia of the tongue is frequently seen in hemiplegic patients The tongue cannot be protruded upon request, but a moment later the patient may without thought lick his lips

Agnosia (not knowing) is a defect higher than the mere inability to perceive tactile, visual, auditory or other forms of sensation, it results rather from the failure to interpret sensory impressions which enable an object, sound, or symbol to be recognized and have meaning "Word blindness" and "word deafness" are forms of visual and auditory agnosia, respectively

Astereognosis is a disorder in which though sensations of touch and muscle sense are retained, the patient cannot recognize an object placed in his hand if his eyes are closed Visual agnosia is seen in lesions of the occipital lobe of the dominant hemisphere, auditory agnosia in injury to the temporal cortex, and astereog-

nosis in lesions of the parietal lobe posterior to the postcentral gyrus

Agraphia or dysgraphia, the inability to write or difficulty in writing, may occur with visual agnosia because of a handicap in the recognition of written words, i e , a word blindness

Learning, Conditioning, and Memory

Learning is defined as a more or less permanent modification of response that occurs as a result of practice, experience or observation It is studied by testing input-output relations just as the connections of a stretch reflex through the spinal cord are studied Both the input and the output are more complex, however, the latter usually falling into the category that is known as behavior Learning has been investigated in literally hundreds of different behavioral situations, with results far too diversified for summary here Nevertheless, many of them can be classified under two major headings as being either classical (type I) or instrumental (type II) conditioning

In *classical* or *pavlovian* conditioning (Pavlov, 1928), the stimulus to which the subject is to be conditioned is always followed by the unconditioned stimulus, perhaps food, or an electrical shock The subject has no control over receiving, avoiding, or escaping the unconditioned stimulus, and the conditioned response when it develops seems to come in spite of rather than by the intention of the subject

Instrumental conditioning, on the other hand, refers to procedures in which the subject's behavior to the initial stimulus determines whether or not the subsequent stimulus is presented Thus, the pressing of a bar after the initial stimulus is presented may bring about a reward such as delivery of food, or prevent a punishment such as an electric shock Conditioning has occurred when the subject has learned to make the appropriate response to the initial stimulus

Classical Conditioning

It was Pavlov (1928) who, in the course of extensive studies on the physiology of the digestive glands, first began a systematic experimental investigation of conditioning phenomena Basically, the experimental procedure for classical conditioning involves the repeated pairing of two stimuli: *a*) an *unconditioned stimulus* (UCS) which at the beginning of the experiment evokes a regular and measurable response called the unconditioned response (UCR), and *b*) a *conditioned stimulus* (CS) which initially is ineffective (neutral) and does not elicit the UCR at the outset of the experiment Following repeated presentations of the conditioned stimulus and the unconditioned stimulus, a response resembling the unconditioned response may be elicited upon presentation of the conditioned stimulus alone Such a response is called the *conditioned response* (CR)

CLASSICAL CONDITIONING EXPERIMENT

In most of the conditioning experiments performed by Pavlov, the secretion of saliva was chosen as the response to be measured A ticking metronome (CS) was presented a few seconds before a small quantity of meat powder was blown into the dog's mouth The food, elicited a certain amount of salivation (UCR) which was measured After a certain number of pairings, the sound of the metronome alone (the food being withheld) elicited the secretion of saliva Pavlov called this response a conditional reflex, meaning that the occurrence of the response to the previously neutral cue was literally conditional upon certain operations, namely, the repeated pairings of the neutral cue (CS) with the unconditioned stimulus The original Pavlovian procedures of classical salivary conditioning were subsequently modified by Bekhterev (1933) He substituted an electrical shock to the hind legs of dogs in place of food as the UCS and studied withdrawal and respiratory responses evoked by the noxious leg shock After repeated pairings of the sound of a metronome (CS) with the electric shock (UCS), the CS evoked a conditioned leg flexion.

ELECTRICAL RECORDINGS

Two major classes of bioelectrical events which have been measured during conditioning are *a*) spontaneous electrical activity of the brain (the electroencephalogram or EEG) which has been observed to vary in frequency, amplitude, and other properties during different phases of conditioning, and *b*) electrical activity evoked by a stimulus (e g , light flash, tone, touch) which may show systematic changes in amplitude and wave shape when

the stimulus is paired with an appropriate UCS The prototype of bioelectrical conditioning is the conditioned blocking of the occipital α rhythm (8 to 12/sec frequency) which was first observed by Durup and Fessard in 1936 These workers discovered that the click of a camera shutter used to photograph α blocking elicited by light sometimes produced attenuation of the α activity Previous testing had shown that the noise of the shutter alone had no effect on the electroencephalogram Study of this form of conditioning then began in many other laboratories, and Jaspar and Shagass (1941) were able to duplicate nearly all the phenomena of classical conditioning (e g , extinction, discrimination, delayed, and trace conditioning)

The latency, magnitude, and duration of evoked potentials in the cerebral cortex have been studied during conditioning with both macro- and microelectrode techniques For example, in a detailed study on cats and monkeys, Galambos and Sheatz (1962) used chronically indwelling electrodes in different brain regions for the study of evoked potentials elicited by clicks and light flashes (the CS) A puff of air to the face was used as the unconditioned stimulus Prior to conditioning trials, the CS was presented many times to habituate the evoked response to the click When presentation of the CS resulted in no further change in the amplitude of evoked potentials, conditioning trials were started with the click regularly followed by the puff of air (UCS) The evoked potential to the click increased in amplitude following a few pairings of the CS and UCS (Fig 9 127)

Cerebral cortical activity has been studied also by the use of microelectrode techniques to monitor single cell discharges during conditioning (Jasper, Ricci, and Doane, 1960, Yoshii and Ogura, 1960). These studies suggest that changes in single cell discharge rates do occur during conditioning, although there is great variability associated with the response In the typical experiment, an exteroceptive light flash (the CS) is paired with an appropriate unconditioned stimulus (e g , shock to paw or sciatic stimulation) and the unconditioned response recorded is either an increase or decrease in the discharge rate of individual neurons. After repeated pairings, the light alone is able to evoke the characteristic discharge rate previously associated only with the unconditioned stimulus Repeated extinction and reconditioning of single cells has been reported to occur with this procedure (Kamikawa et al , 1964)

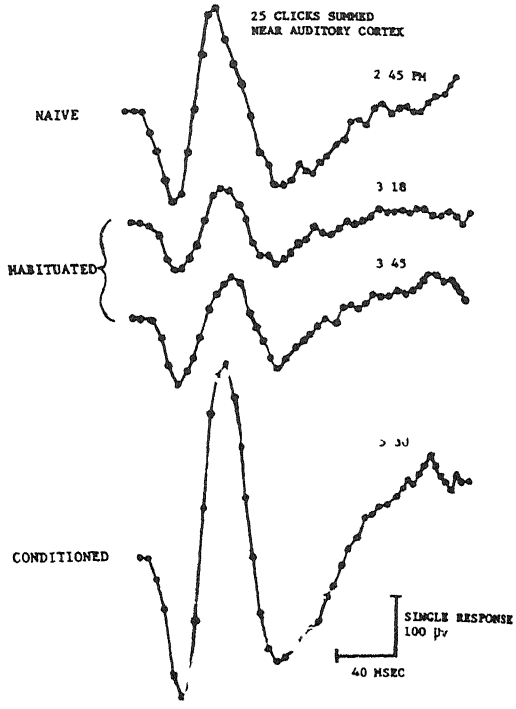


Fig 9 127 Cortical potential records averaged by computer to show response of auditory cortex to sound of a click *Naive* response before habituation *Habituated* response after repeated stimulation *Conditioned* response to click plus puff of air to face (Galambos and Sheatz 1962)

ELECTRICAL STIMULATION

Another technique has been the use of direct brain stimulation as either the conditioned or unconditioned stimulus As might be expected, interruption of the afferent fibers of an exteroceptive conditioned stimulus prevents the acquisition and the performance of a conditioned response However, direct electrical stimulation of the spinal cord, cerebellum, and various areas of the cerebral cortex may be substituted for normal exteroceptive stimuli, and serve as conditioned stimuli For example, instead of pairing an exteroceptive light (the CS) with food (the UCS), electrical stimulation of the occipital region may be paired with the UCS With this procedure, conditioning takes place in about the same number of trials as when normal exteroceptive signals are used as

conditioned stimuli Even electrical stimulation of nonsensory areas of the cortex has been used successfully as a conditioned stimulus It has also been possible to obtain successful conditioning where both the conditioned and unconditioned stimuli are applied to the cortex electrically (Doty, 1961)

Instrumental Conditioning

If the response of the animal can affect the occurrence or duration of the UCS, then the procedure becomes one of instrumental conditioning The most common instrumental procedures used in the study of the neurophysiology of behavior may be classified as *a*) escape and avoidance conditioning, *b*) operant conditioning, *c*) discrimination learning, and *e*) problem-solving learning The same parameters of conditioning (generalization, extinction, etc) discussed for classical conditioning procedures are also found in instrumental conditioning procedures

ESCAPE AND AVOIDANCE CONDITIONING

In *escape* conditioning, the animal can terminate an aversive stimulus by making an appropriate response In *avoidance* conditioning, the response prevents the delivery of the aversive stimulus The most common responses studied have been paw-withdrawal, jumping, lever pressing, wheel-turning, and shuttling In this last procedure, the animal is placed in a two-chambered box with the two compartments separated by a hurdle, a swinging door, or a line on the floor In order to avoid shock, the animal has to learn to respond to a signal (the CS) by shuttling or hurdling from one compartment into the other compartment This is an example of an *active* task in which the animal must move or make some other response to avoid the UCS Other procedures, where the animal must inhibit a previously learned response to avoid an aversive stimulus, are cases of *passive* avoidance For example, in passive avoidance an animal may be tained to enter a compartment or to insert its head into a tray for food when a CS is presented After the animal has learned this response, an aversive stimulus is then presented whenever the animal enters the compartment or inserts its head into the tray when the CS is presented The animal, in order to avoid the shock, must now learn to withhold his approach to food The behavior of not approaching food is passive avoidance, whereas an animal that must move or make some active response (turning a wheel) is displaying active avoidance

OPERANT CONDITIONING

The development of operant conditioning procedures is closely associated with the experimental analysis of behavior by B F Skinner (1938) and his associates In the typical "Skinner box" situation, a hungry rat is placed in a dimly lit box which has a movable lever connected to an electric switch The rat explores the box and occasionally by chance presses its paws upon the lever The number of presses for some arbitrary time period defines the rat's *preconditioning operant level* of bar pressing The food magazine is then connected to the switch so that when the rat presses the bar a pellet of food falls into a dish The rat eats, and soon presses the bar again The delivery of foot pellets (the click of the food magazine, the food dropping into the food tray, etc) "reinforces" bar pressing The number of lever presses plotted against time provides a record of the course of operant conditioning (or operant rate) Often, the untrained rat is *shaped* to perform the desired bar pressing response by the procedure of *successive approximations* For example, food is delivered whenever the hungry rat approaches the bar, then whenever it touches the bar, and finally only after it presses the bar

In operant conditioning the animal must make a response before the response can be reinforced Skinner (1938) called this behavior "operant" because the behavior of the rat operates on the environment The operant conditioning situation differs from both classical and avoidance conditioning in that *a*) there is no specifiable unconditioned stimulus which elicits the response, and *b*) the animal is free to respond at any time in the operant situation, but is usually given discrete trials in avoidance and classical conditioning Operant techniques have been used in the experimental analysis of the behavior of a wide variety of animals including humans. The objective of operant techniques is to place some arbitrary sample of behavior under experimental control so that behavioral processes may be investigated as a function of a wide variety of conditions

Reinforcement: Definition and Parameters. The term reinforcement is commonly used when discussing operant conditioning, although the pairing of the conditioned and unconditioned stimulus in classical conditioning is also called reinforcement A reinforcing stimulus may be operationally defined as a stimulus which alters the probability of occurrence of the response that precedes it Reinforcing stimuli (reinforcers) may be either positive (e g , food, water) or aversive (e g , electric shock) The quantity and quality of the reinforcement as well as the interval between the response and the presentation of the reinforcement (called delay of reinforcement) are important variables which affect the organism's behavior during learning (cf Kimble, 1961) Another important variable is the schedule of reinforcement Reinforcement may be given on every trial or after every correct response (*continuous reinforcement*) or intermittently, that is, the response is reinforced only part of the time The main effect of such an *intermittent* or *partial reinforcement* is to make it more difficult to extinguish the response once it has been learned The effect of intermittent reinforcement is found in both operant situations and discrete trial instrumental tasks Partial reinforcement in the operant situation may be delivered on either a temporal (interval) schedule or on a response (ratio) schedule Further details about these techniques should be looked for elsewhere, for example, in Kimble (1961) or in textbooks of psychology

Electrical Self-Stimulation of the Brain. The discovery that low voltage electrical stimulation of various brain loci may be used as positive and negative reinforcement has led to numerous self-stimulation studies In these studies, stimulating electrodes are chronically inserted into subcortical structures and connected by a light flexible cord to a source of stimulating current The animal, by pressing a bar or lever, delivers a flow of stimulating current to its own brain The electrical circuit is constructed so that a fixed duration of stimulating current is delivered for each bar press Thus, the experimental procedure is the same as other operant techniques, except that instead of receiving a food pellet as reinforcement, the animal receives a slight shock to its brain

Olds and Milner (1954) were the first to discover that rats, allowed to stimulate their own brains, would maintain high lever pressing rates over long periods of time without any other reward This interesting and important finding has resulted in extensive experimentation and speculation concerning the variables which influence brain self-stimulation Some of the variables shown to be critical determinants of this phenomenon are the specific anatomical locus of the electrode, the intensity of the stimulating current, the schedule of brain reinforcements, and food and water deprivation For example, rats with electrodes in the lateral hypothalamus will, under continuous reinforcement, press about 10,000 times an hour for brain stimulation but only about 500 times an hour when the electrodes are in the septal or amygdaloid areas With electrodes in the periventricular system of fibers (dorsal longitudinal fasciculus of Schutz), rats will actively work to prevent or escape the brain stimulation Stimulation of certain medial hypothalamic areas yields both positive and negative effects With electrodes in these brain sites animals can be trained to turn on the brain stimulation and then immediately turn it off (cf Olds, 1962) The rewarding effects of intracranial self-stimulation have also been observed in cats, monkeys, and humans (Sheer, 1961) During neurosurgery of humans, stimulation of various brain loci has resulted in verbal reports of feelings of joy and satisfaction and experiences of tickling and fluttering of muscles in the pelvic region which were regarded as pleasurable The patients, given a key which controlled the delivery of the brain stimulus, would eagerly respond by self-stimulation

DISCRIMINATION LEARNING

Discrimination learning is involved in every learning situation, but has been studied most extensively with instrumental procedures The general procedure used to establish a discrimination involves the extinction of responses to certain stimuli by withholding reinforcement, while a response to some second stimulus is maintained by the delivery of reinforcement The discrimination experiment was originally designed to test the sensory capacities of various animals At present, discrimination behavior is often studied as a learning phenom-

enon in its own right The two methods of presenting stimuli in discrimination experiments are *a*) successive presentation of stimuli (contrasts), and *b*) simultaneous presentation of stimuli (choice response)

Types of Discrimination Task Visual, somesthetic, and auditory discriminations have been studied in a variety of instrumental conditioning procedures The most intensively studied of all types of discrimination has been visual discrimination—particularly form discrimination In the typical discrimination task with primates, two stimulus objects which may differ in either shape, size, or color are placed over two food wells on a test tray (Fig 9 128) The food well under the correct object is baited with food and the other food well is left empty The animal must learn to select the correct stimulus and move it aside to pick up the food in the well The positions of the objects on the tray are shifted from one side to the other in either a random or balanced order Between trials, a screen is lowered in front of the monkey so that the appropriate food well may be baited Information about these and other procedures used in behavioral testing of primates may be found in the volume edited by Schrier et al (1965)

PROBLEM-SOLVING PROCEDURES

Problem-solving instrumental procedures are no doubt the most familiar of the animal testing procedures developed by psychologists Originally, many of the so-called problem-solving tasks were designed to test for symbolic behavior in lower animals Now, it is generally assumed that the animals learn problem-solving tasks in the same way as other instrumental tasks are learned, except that the discriminative cues are more subtle or complex in the problem-solving tasks Problem-solving procedures have been classified as learning with complex cues (cf Ratner and Denny, 1964) In general, the following types of learning tasks may be distinguished

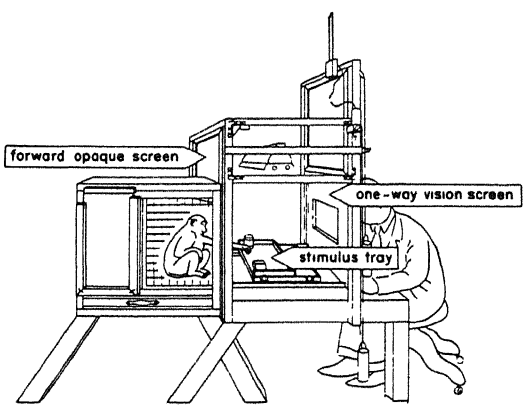


Fig 9 128 General test apparatus for studying learning by way of discrimination performance In this form it is known as the Wisconsin or the Harlow apparatus

Maze Learning. Discrimination learning in either a T-maze or a Y-maze is one of the most widely used instrumental learning techniques The discriminative cue may be either spatial or nonspatial In a spatial problem, the animal chooses between a pair of stimuli which differ only in their position in space The animal, for example, may be required always to turn left or choose the left door in order to receive reinforcement In a nonspatial problem, the two discriminative cues each appear in the two positions in a random order, and reinforcement is correlated with the positive cue For example, in a red-green visual discrimination, the green alternative may be reinforced independently of its position Spatial and nonspatial problems may also be used in complex mazes For example, at each choice point the animal may be required always to turn left (spatial) or to turn into the lighted alley in a light-dark nonspatial discrimination problem

In some cases, the animal's own response may be the cue for the correct response at the next choice point For example, in a *simple-alternation* problem, the animal must alternate left and right turns (LRLR) in a 4-choice maze Turning left at the first choice point provides the cue for the right turn at the next choice point and so on Most mammals can readily learn a simple alternation problem as they tend normally to alternate spontaneously in most maze situations. A more difficult problem is the *double-alternation* pattern (LLRR) Rats, typically, fail to learn a double-alternation problem, whereas other mammals (cats, dogs) are capable of learning a LLRR problem but fail to learn a longer sequence

Delayed Response. The delayed-response problem is one of the most sensitive behavioral tests and has been used in neurobehavioral research to detect the effects of brain lesions, drugs, and radiation The delayed-response

problem is particularly sensitive to damage to the central nervous system, especially the frontal lobes In this problem, the animal is presented a limited number of choices where food might be placed The animal is first trained to respond to a particular stimulus in order to obtain food The animal is then restrained from approaching the food compartments or the stimuli over the food wells until released The experimenter at the beginning of the trial permits the animal to see where the food is located by turning on a light over the correct food compartment (indirect method) or by directly placing food into a particular food well (direct method) The light is then turned off or a screen is lowered in front of the animal, and following a certain delay, the animal is allowed access to the compartments or food wells The purpose of the testing procedure is to determine how long the animal can be delayed without "forgetting" where the food has been placed This task commonly has been used to determine if an area of the brain is involved in immediate memory

MECHANISM OF CONDITIONING

One of the oldest theoretical controversies in learning theory concerns the question of just how does conditioning occur Pavlov believed and many authors after him assumed that classical conditioning is a neocortical phenomenon It was supposed to be the only level of the nervous system with the plasticity necessary for development of what appear to be new connections between stimulus and response Later experiments showed, however, that conditioning can be induced in decorticate animals (Hilgard and Marquis, 1940), and in spinal preparations (Dykman and Shurrager, 1956), albeit only after prolonged training using intense stimuli Still more recently single neurons have been found to exhibit responses that are considered to be conditioned (Tauc, 1967) Consequently, the capacity for conditioning is now regarded as a generalized property of neurons, not necessarily dependent upon the organization of a higher nervous system

Another point of uncertainty has been the relation of somatic muscular movement to conditioning, because it was supposed that a feedback to the nervous system from skeletal muscles was needed for the conditioning process The truth of this concept has been tested through the use of drugs, e g , curare, that block somatic movement Conditioning does occur in animals under the influence of curare, and the conditioning persists into the postcurare period (Black et al , 1962, Solomon and Turner, 1962) In fact, Miller (1969) has reported that "learning" was actually facilitated by administration of curare He suggested that this was because the drug removed " interference from skeletal responses and distraction from irrelevant cues" (p 444)

AUTONOMIC CONDITIONING

A closely related question is whether conditioning can take place in the autonomic nervous system, independently of somatic motor response One of the principal problems has been the exclusion of somatic movements that might have produced secondarily what appeared to be a conditioned autonomic response For example, although heart rate is completely under autonomic control and should be a good subject for a test of conditioning, it is modified also by changes in the respiratory pattern that are brought about through somatic nerves What is clearly autonomic conditioning, however, has been demonstrated by Miller (1969) and his associates Using instrumental techniques they observed in rats the conditioning of salivation, heart rate, intestinal contraction, glomerular filtration, renal blood flow, and the rate of urine formation (Fig 9 129) By appropriate reinforcement they could either increase or decrease any of these variables Their experiments, therefore, confirm the conclusion that conditioning is a much more common or generalized phenomenon than most of the earlier authors had supposed

NEURONAL CONDITIONING

In higher animals because of the multiplicity of neurons and pathways involved in learning and/or conditioning, it has been extremely difficult to discover cellular correlates of the conditioning process Lower animals, by contrast, have simpler nervous systems and lend themselves to study of fundamental processes in single cells The discovery that these lower forms exhibit conditioning has revolutionized this field of study In the gastropod, *Aplysia*, the neurons of the abdominal ganglion are large, they are practically uniform in appearance and location from one animal to another,

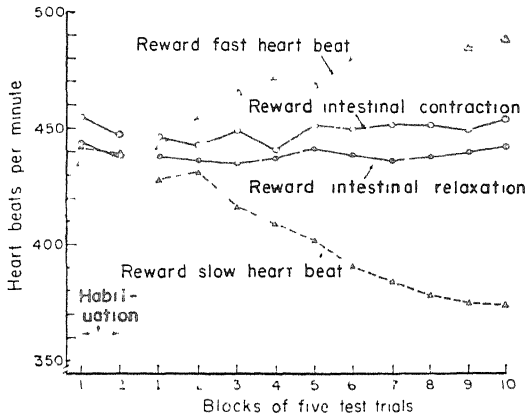


Fig 9 129 Learning curves for rats in which heart rate was recorded during experiments where reward was provided by stimulation of the brain All records portray rate of heart beat but the physiological change for which reward was given was different for each curve The two middle records show that rewarding changes in intestinal contraction did not alter heart rate Miller and Banuazizi 1968 cited from Miller 1969

they have identified connections, and can be penetrated under microscopic observation with microelectrodes for recording of transmembrane potentials, for intracellular stimulation or for microinjection Using these techniques the properties of their membranes have been described, synaptic potentials have been recorded, and the role of chemical transmitters has been clarified (Tauc, 1967) With reference to these properties these neurons are not unique, but rather serve as convenient examples of what happens in nerve cells upon stimulation Their value in the study of conditioning, however, is derived from the disclosure that their sensitivity to stimulation is a function of their immediate history (Tauc, 1967, Kandel and Spencer, 1968) Whether or not a cell will respond to a stimulus of a given strength depends upon whether or not it has been previously stimulated, and the conditions of the previous stimulation Furthermore, when stimuli are delivered via two different afferent nerves they interact at the synaptic membrane of the ganglion cell in a fashion that appears to simulate pavlovian conditioning These and related observations have given a strong impetus to theories in which conditioning and learning are regarded as localized changes in or on the surface of neurons Although there seems to be no doubt that in higher nervous systems and in more complex responses the phenomena of conditioning, learning, and memory involve pathways and circuits rather than single cells, yet what happens in a circuit must be determined ultimately by the state of the cells that comprise it One may suppose that *learning* is characterized by physicochemical changes in cells as a result of *conditioning*, and that these changes in turn create the possibility of *memory*

Memory

The exact nature of the changes in the nervous system that constitute the memory trace remains unknown There are, however, a number of studies that shed some light on this fascinating problem (See John, 1967, Byrne, ed , 1970)

One of the early speculations concerning the establishment of the memory trace is generally attributed to Muller and Pilzecker (1900) In order to account for the phenomenon of decreased retention of recently learned material when another activity intervenes between initial learning and retention testing, Muller and Pilzecker proposed the existence of a neural perseverative process This perseverative process was presumed to be easily interfered with by external influences and was also considered necessary to the "consolidation" of the memory trace for recently learned material Most of the more recent investigators have accepted this concept, namely, that the establishment of a memory trace takes place in two stages, an initial stage in which the memory trace is evanescent and easily disrupted, and a later, more stable state, dependent on the initial process and which is the "permanent" memory trace

For many years the only evidence in support of such a proposal was the retrograde amnesia seen in many patients after cerebral trauma Head injuries are often followed by loss of memory for events occurring during the period preceding the injury With the advent of electroshock therapy for psychoses in the late 1930's, further evidence for this two-stage nature became available Shock is administered through electrodes applied to the skin of the head It leads to unconsciousness, and presumably disrupts by electrical energy much if not all of the neuronal circuits of the brain—at least temporarily Many physicians and investigators who used or studied electroconvulsive shock (ECS) found that it too produces a dif-

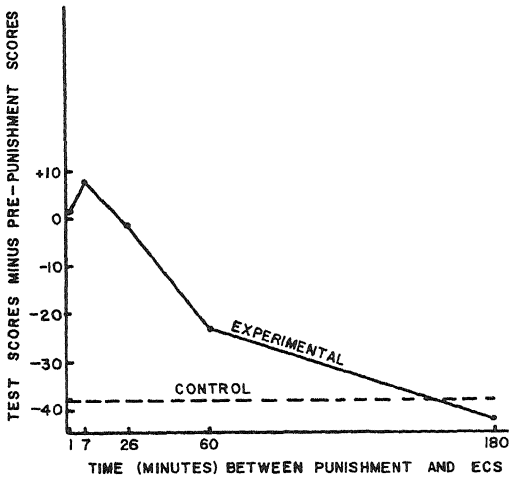


Fig 9 130 Effect of electroconvulsive shock (ECS) on retention as a function of the time interval between the learning experience and the application of ECS The data show that if the interval between learning and ECS is as long as 180 min the ECS does not prevent learning retention (Heriot and Coleman 1962)

ferential loss of memory only for events immediately preceding the shock treatment (e g Williams, 1950) The time course of the effects of ECS on the memory trace has since been explored in a number of studies For example, Heriot and Coleman (1962) trained rats to press a lever for a food reward Then the lever was electrified so that the animals received a slight shock when the lever was touched After this, animals avoided the lever The animals were divided into groups which received also an ECS at intervals ranging from 1 to 180 min after the initial, slight shock Control animals and animals receiving ECS 180 min after being shocked by the lever showed a significant decrease in number of lever pressings Animals receiving ECS within an interval of 26 min pressed the lever at the same rate as before being shocked, they apparently did not remember their unpleasant experience with the lever (Fig 9 130) Other studies have shown that this consolidation phase may require from about 15 min to possibly several months (Gellhorn, 1945), depending on the learning situation and on the procedure used in disrupting the consolidation Other procedures that have been successfully utilized to disrupt the consolidation phase include hypothermia, insulin coma, Metrazol convulsions, anoxia, some anesthetic agents and application of chemicals or electrical stimulation to local regions of the brain

NATURE OF THE STAGES

If the sequence of conditioning-learning-memory is analogous to other types of neural activity, then it must depend ultimately upon conduct of nerve impulses through circuits of neurons Perhaps the simplest way in which this can be described is to suppose that when a sensory field, for example, is registered upon neurons of the cerebral cortex (phase I, Fig 9 131), the circuit through which the activity was carried from brainstem to cortex becomes temporarily self-exciting, because of changes of the type exhibited by *Aplysia* ganglion cells, as described above (phase II, Fig 9 131) This activity may or may not become evident at the level that is called consciousness, but because the circuit remains in action its activity can be diverted to the conscious level rather easily where it becomes evident as memory During the time that the sensory information is stored in such a circuit it can be either retrieved or it can be destroyed, as in the case of ECS treatment This is the initial stage of learning, the stage of "registration" (Jarvik, 1970) Inasmuch as all neural activity is based upon physicochemical changes both along axons and at synapses, this stage must have such a basis

If the activity in the circuit is intense enough, or if it persists long enough, it is assumed to create more nearly permanent changes in the neurons involved (phase III, Fig 9 131) As a result of these changes any of the cells that carried the activity is likely to re-establish the complete circuit whenever it is activated This means that a circuit that in the first instance was integrated into a pattern because of repetitive neural activity has now been altered so that the pattern tends to be reassembled because of some structural or

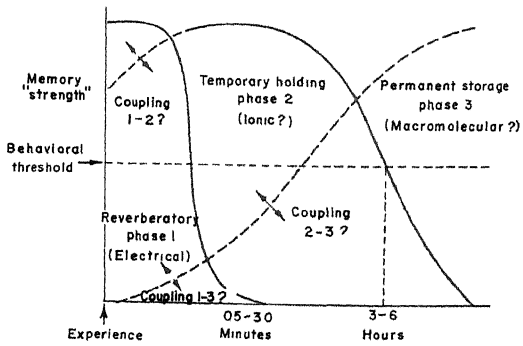


Fig 9 131 Relation of time to "strength" of memory, with curves showing hypothetical duration of three phases in memory process (John, 1970)

chemical change in its elements This is the stage of "retention," brought about by "consolidation" The nature of the alterations that convert phase II into phase III has been studied in many laboratories, with results that are fascinating even though as yet they are not entirely consistent in their significance

CHEMICAL CORRELATES OF MEMORY

There are many indications that intracellular protein synthesis involving ribonucleic acid (RNA) may play a vital role in the long-term memory trace RNA is a substance capable of preserving a structural modification by continually imposing on molecules being formed with neural tissue a structure that could be determined by learning and experience One of the first to suggest that RNA might be this substance was Hyden (see 1962), who showed that stimulation of the brain tissues of rats and rabbits caused an increase of RNA in the neurons In later studies Hyden dissected out, and chemically analyzed, single cells from the lateral vestibular nucleus of rats trained to climb and balance on a steel wire inclined at a 45-degree slope, in order to obtain food He found that significant changes took place in the composition of the nuclear RNA Of the four amino acids that form RNA—adenine, cytosine, guanine, and uracil—Hyden found an increase in adenine and a fall in cytosine

More recently there have appeared reports of studies which suggest that RNA altered by learning may transfer the memory trace from a trained animal to a naive animal One group (Babich et al , 1965) trained one group of rats to press a lever for food in a Skinner box only when a blinking light was on Another group of rats was trained to press a lever for food only when a click was sounded RNA was then extracted from the brains (excluding cerebellum, brainstem, olfactory bulbs, and prefrontal areas) of these animals and injected intraperitoneally into untrained rats Some of the untrained rats received RNA from light-trained rats, other received RNA from click-trained rats The untrained rats were then tested in a Skinner box for number of lever presses, with both click and light flash (at separate times) The click-injected group responded more to click than to light, and the light-injected group responded more to light than to click

Dingman and Sporn (1961) studied the effects of 8-azaguanine on learning in rats in a water maze 8-azaguanine is so similar to the guanine component of RNA that enzymes responsible for RNA synthesis will employ the 8-azaguanine rather than guanine and proceed to manufacture an RNA quite unlike the usual product The administration of 8-azaguanine did not affect the ability of rats to swim through a maze learned some time beforehand, that is, it did not affect recall and performance of a well-established memory trace Dingman and Sporn next injected rats with 8-azaguanine and then tried to teach them to swim a new maze The injected rats made twice as many errors as a control group, suggesting that the formation of new memory traces is depressed when RNA synthesis is interfered with

Research on this problem, however, has become one of the most controversial topics in neurophysiology, in part because the results reported have seemed inherently implausible to many investigators, and in part because of reports that certain of the results could not be verified The literature is much too large for evaluation here, further information is available in the proceedings of a recent conference on the subject (Byrne, editor, 1970)

Another means of studying the chemical basis of memory is to alter the synthesis of the protein that is guided by the cellular RNA Thus Flexner et al (1963) reported that previously stablished memory traces are disrupted when animals are given injections which may suppress protein synthesis by neuronal RNA Mice were trained to run to one arm of a Y-maze At various times after training, the animals were given local intracerebral injections of puromycin, an antibiotic which inhibits protein synthesis in vivo The intracerebral injections were administered bilaterally in either the temporal region, the ventricular region, or the frontal region and all combinations of these sites Short-term memory was lost only with the temporal injections Loss of longer-term memory required puromycin over basically the entire cortex and possibly the thalamus as well

The results of these experiments seemed clearcut when they were first presented, but subsequent work has altered some of the interpretations originally offered (Dethier and Stellar, 1970) The principal complication is that puromycin has effects other than the inhibition of protein synthesis, one of which is that it al-

ters cerebral electrical activity (see Barondes, 1970). The reason for this alteration is not known, but it apparently is not a direct consequence of the blocking of protein synthesis, inasmuch as other blocking agents do not lead to these electrical changes. Puromycin is assumed to evoke a subconclusive electrical activity in the cortex, which in itself might have an effect on memory consolidation. Nevertheless, another blocking agent, acetoxycycloheximide, likewise interferes with retention of learned behavior without altering the cortical EEG. This supports the original conclusion that protein synthesis is required for consolidation (Barondes, 1970; Agranoff, 1970). The data currently available are in agreement with the hypothesis that repetitive activity in a neuronal circuit leads to changes in RNA composition, as a result of which protein synthesis is altered in a fashion that permanently establishes the possibility of reactivation of that particular circuit.

No studies have delimited any single region of the brain as *the* site of long-term memory. Many studies have shown deficits in the retention of specific tasks after extirpation of localized portions of cortex or subcortical structures without any generalized deficit of long-term memory per se. For example, a learned auditory intensity discrimination task is lost after removal of auditory cortex (Raab and Ades, 1946). However, tasks involving other modalities are not lost, and the auditory intensity discrimination can be relearned. It seems, although the capacities necessary for performance of the task may have resided in the cortex in the intact animal, subcortical centers were capable of providing the requisite capacities for learning the discrimination after the cortical removal (as noted above). Current evidence, then, indicates that the capacity for long-term memory is widely diffused throughout the nervous system with the memory trace stored in multiple locations in the intact animal.

DISORDERS OF NEURAL CONTROL

It is obvious that lesions at various points in the nervous system are likely to produce different manifestations These differences may be revealed as disorders in somatic muscular activity Many such disorders have been referred to in preceding chapters because of the insight they provide into normal functions of the several parts of the brain and spinal cord This chapter is devoted to brief consideration of selected but typical clinical syndromes, presented in accordance with the concept of levels of function as in foregoing pages

Significance of the Level of a Lesion

Destruction of the peripheral nerves supplying a skeletal muscle produces an easily understood clinical picture The muscle is *1)* totally paralyzed, it is beyond the patient's control, and he cannot move it, *2)* no reflex activity of any sort is possible, although in the early stages the muscle may still contract in response to direct percussion or direct stimulation with a galvanic current, and *3)* the muscle wastes away to a useless fibrous cord because it is devoid of innervation and completely without use Involvement of this sort is often spoken of as a "lower motor neuron" paralysis and it may develop following destruction of the anterior horn cells in the spinal cord, or following division of the anterior spinal roots or of the muscular nerve One must never forget, however, that not all paralyses are complete Some innervation may remain and then the muscle is weakened but not completely paralyzed, the reflexes are diminished but not abolished and the wasting is incomplete

Although the term "lower motor neuron" paralysis is accurate for the condition described above, it has the disadvantage that it implies that there is a counterpart, an "upper motor neuron" paralysis This is misleading for there is no constant reliable entity which can be so designated The manifestations of involvement of the many neurons which lie above the anterior horn cell which are responsible in one way or another for some phase of muscular activity vary widely depending upon which such neurons or groups of neurons are involved Thus, a severe paralysis with spasticity and hyperactivity of the stretch reflexes may occur if the descending motor pathways in the internal capsule are involved If, on the other hand, the lesion is limited to the corticospinal fibers in the central portion of the cerebral peduncle one sees a loss of some fineness and delicacy of movement with little or no spasticity and little increase in tendon reflexes Lesions of the globus pallidus combined with destruction of the substantia nigra are characterized by increased resistance to passive manipulation of the extremities, with a generalized rigidity of the body and a loss of the facial movements of expression and of many automatic associated muscular movements but with little or no true paralysis and little if any change in the tendon reflexes As noted before a paralysis arising from a lesion in the internal capsule is commonly associated with a flexor spasticity in the upper extremity and an extensor spasticity in the lower extremity, whereas a severe lesion in the mesencephalon is likely to produce a paralysis with extensor spasticity in all four extremities In contrast to this, a transverse lesion of the spinal cord will produce a complete paralysis of voluntary movement in the skeletal musculature below the level of the lesion, with a variable reflex pattern, depending upon the location of the lesion and the pathways which are involved One reflex seems to have its own specific significance—the *sign of Babinski* (see below) This is a sign of interruption of the corticospinal fibers which innervate the musculature of the great toe, it is not a sign of damage to any other corticospinal fibers

Both "upper" and "lower" motor neuron paralyses may be associated with atrophy or wasting of the involved muscles Muscular wasting is in general proportionate to the degree of loss of use of the muscle rather than to

loss of innervation Wasting from "upper motor neuron" lesions is never complete because even with a *total* paralysis of voluntary activity some reflex activity persists

Lesions at various levels are likely to involve other structures, and thus indicate their location by the nature of the other clinical manifestations which the patient develops (Fig 9 132) Thus, a massive lesion of the internal capsule will produce a contralateral hemihypesthesia and possibly a homonymous hemianopia as well as a hemiplegia A unilateral lesion in the midbrain often will produce a third cranial nerve palsy (with a ptosis of the upper eyelid, a paralysis of several ocular muscles, a dilatation of the pupil and a loss of the pupillary light reflex) on the side of involvement in addition to a contralateral spastic hemiplegia Lesions involving one half of the lower brainstem may be associated with a paralysis of one or more cranial nerves (e g , facial, vagus or hypoglossal) on that side of the body and a hemiplegia and hemianesthesia on the opposite side of the body In contrast with this a unilateral lesion of the spinal cord produces what is known as a Brown-Sequard syndrome—a paralysis of the skeletal muscles of the trunk and extremities and a loss of position and vibratory sensibilities below the level of the lesion and on the same side, with a loss of pain and thermal sensibilities on the opposite side (Fig 9 133) Obviously the possible combinations of symptoms and signs are numerous and cannot all be discussed in detail here Their recognition and proper interpretation are dependent upon a thorough understanding of the structure and function of the nervous system

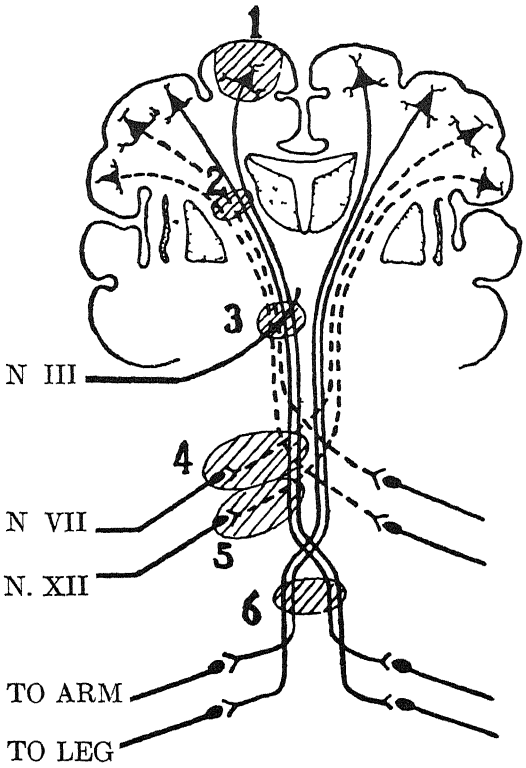


Fig 9 132 Diagram illustrating the clinical signs that follow typical lesions of the central nervous system *1)* Lesion of cerebral cortex causes monoplegia or hemiplegia depending upon its extent 2) Lesion in internal capsule may cause hemiplegia *3)* Lesion of midbrain may involve descending motor fibers to opposite side of body and an ipsilateral eye muscle nucleus *4)* or *5)* Lesion in pons or medulla has results similar to location *3,* but involving other cranial nerves *6)* Transection of spinal cord causes paraplegia

It is evident that the paralysis is more likely to be bilateral if the lesion is situated in a region where, as in the brainstem or spinal cord, the descending motor tracts of the two sides are approximated, than if it is located at a level where these tracts are widely separated, as in the corona radiata or internal capsule, then unilateral effects are usual Complete interruption of the spinal cord will of course result in paralysis on both sides of the body below the level of the lesions—*paraplegia* (see next paragraph)

Lesions of the Spinal Cord

TRANSVERSE DIVISION OF THE SPINAL CORD

A sudden, or rapidly progressive, complete interruption of the continuity of the cord may result from injury (e g , gunshot wound, fracture-dislocation of the spine, etc) or from acute inflammation (e g , transverse myelitis) Immediate and complete loss of voluntary power below the level of the lesion results Paralysis of both lower limbs resulting from this or any other nervous lesion is spoken of as *paraplegia* Complete division of the cord in the lower cervical region will result in paralysis of all four limbs, *quadriplegia* A lesion of this nature in the upper cervical region is of course rapidly fatal since the diaphragm and other respiratory muscles are isolated from the respiratory center

Stage of Spinal Shock Immediately fol-

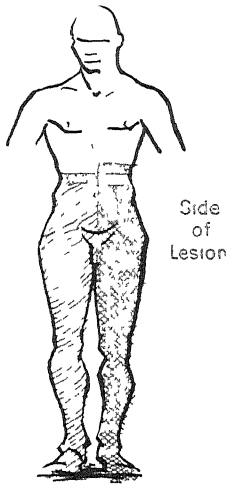


Fig 9 133 Illustration of motor and sensory deficits following hemisection of the lower thoracic spinal cord Narrow white band at level of lesion indicates total anesthesia and analgesia ▨ Analgesia thermo anesthesia and partial loss of tactile sense ☐ Cutaneous hyperthesia ■ Motor and vasomotor paralysis loss of joint muscle vibratory sense

lowing the injury there is complete loss of visceral and somatic sensations, and flaccid paralysis below the level of the lesion The skeletal muscles are quite toneless The tendon jerks, plantar response, and abdominal reflexes are abolished The cremasteric and bulbocavernosus reflexes, though absent as a rule, may at times be elicited The anal reflex is present A zone of heightened sensitivity (hyperesthesia) immediately above the level of the lesion may be present and spontaneous pains in this region, or a feeling of tightness encircling the body may be experienced by the patient There is retention of urine and feces This stage is analogous to the state of spinal shock in lower animals but is much more severe and prolonged (chap 5) In man spinal shock lasts for from 1 to 3 weeks

Stage of Reflex Activity. After a variable period of time, depending largely upon the general health of the patient, reflex activity begins to return to the isolated spinal cord below the level of the lesion. The first reflex to appear is usually the sign of Babinski, the extensor response of the great toe to stimulation of the sole of the foot. This reflex is actually a part of the flexor reflex pattern of the lower extremities Both exaggerated extensor and flexor reflex patterns are usually present with lesions of the descending motor pathways They vary in intensity and in relation to each other If the lesion lies above the pons the extensor reflexes usually predominate If the lesion is in the spinal cord the flexor reflexes are usually more prominent than when the lesion lies at higher levels and they may exceed the extensor reflexes in intensity However, this balance between extensor and flexor reflexes fluctuates from time to time in any patient and is quite variable from one patient to another After the sign of Babinski has become elicitable other flexor reflexes, dorsiflexion of the foot, flexion of the knee and hip gradually become more marked and more easily elicited At first they are usually seen only with stimuli which except for the interruption of the pain pathways in the spinal cord would be painful, i e , nociceptive stimuli Later less intense stimuli such as passive flexion of the toes or mere touching of the foot or leg may be sufficient to provoke a violent flexion of the leg When this stage is reached the patient usually has a *paraplegia in flexion* The extensor reflexes begin to return after variable periods At first the only demonstrable reflex of this type may be the patellar tendon reflex (knee jerk) Later the Achilles tendon reflex (ankle jerk) will return, ankle and patellar clonus may be elicited At times the patient may develop severe extensor spasms and the muscles may be involved by severe clonic contractions with little or no external stimulation At other times the extremities may lie quietly without either extensor or flexor reflexes causing any striking disturbances Elicitation of the flexor reflexes will immediately result in a subsidence of ankle clonus or other manifestations of the extensor reflexes, and vice versa This inhibition of the one when the other is produced is due to reciprocal innervation On occasion elicitation of the flexor reflexes will be associated with evacuation of the bowel and bladder However, reflex or automatic evacuation of the bowel and bladder need not be associated with a violent flexion of both lower extremities Such reflex evacuation may result from distention of the bladder or rectum as the case may be, or may be induced by the patient He may produce emptying of the bladder by stimulating the skin or the inner aspect of the upper thigh or reflex defecation by stimulating the rectal mucosa at and just above the anal ring These reflexes are often used to great advantage by the paraplegic patient The time of

development of these various reflexes and their vigor depend in large measure upon the condition of the patient and his paralyzed extremities A poor state of nutrition or the presence of infection in the urinary tract or in decubitus ulcers, the development of severe muscular wasting or of fibrous changes in the muscles, or about the joints will all mitigate against the development of active reflex activity

INCOMPLETE INTERRUPTION OF THE CORD

A bilateral lesion of gradual onset which impairs the activity of the descending motor pathways usually causes *paraplegia in extension* The muscles are spastic and extensor activity predominates The tendon jerks are exaggerated and patellar and ankle clonus can be readily elicited Paraplegia in extension is more or less comparable to decerebrate extensor rigidity The "extensor" plantar response, which is in reality a part of the flexor reflex pattern, is present, but much less intense than in paraplegia in flexion It is associated with much less contraction in other flexor muscles and may be accompanied by a crossed extensor reflex The mass reflex is absent The abdominal reflexes are lost

The Plantar Reflexes. This is the most appropriate place to consider these and other superficial reflexes

The *normal plantar response* to a light scratch applied to the skin of the sole is plantar flexion of the four outer toes with no movement, or, more usually, plantar flexion of the great toe The center for the reflex lies in the first sacral segment, its physiological significance is unknown With a lesion of the corticospinal fibers from the precentral region to the foot at any level above the first sacral segment, the normal response is replaced by one in which *dorsiflexion* of the great toe and often spreading or fanning of the outer toes occurs This response is called after its discoverer the *sign of Babinski* (Fig 9 134) It is elicited best from the outer border of the sole The upward movement of the great toe is part of a general flexor response homologous with the flexor reflex elicitable from the hind limb of a lower animal (p 9–70) The flexor nature of the reflex is shown by the associated contraction of the hamstrings, i e , semitendinosus, semimembranosus, and biceps femoris, which usually occurs Its flexor nature is also indicated by the fact that an undoubted extensor reflex such as

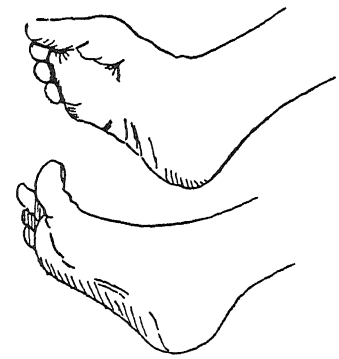


Fig 9 134 *A* Normal response to plantar sensory stimulation *B* Sign of Babinski following similar stimulation

ankle clonus is readily inhibited by evoking the Babinski reaction (reciprocal inhibition)

An "extensor" plantar response is present normally in infants up to the first year or so, i e , to the age of walking, and is then probably due to the undeveloped state of the corticospinal tracts It is also present in normal adults during deep sleep and in the apneic stage of Cheyne-Stokes respiration, being apparently due in the latter instance, to anoxemia of the motor cortex

Other Superficial Reflexes. *1) Oppenheim's reflex* is simply a modified Babinski response, it is associated with the same conditions as the latter and has a similar significance It consists in dorsiflexion of the hallux which results when a firm downward sliding pressure is applied to the skin over the anterior border of the tibia The Gordon, Chaddock, and Gonda reflexes are other modifications of the Babinski response

2) Abdominal reflexes Lightly scratching the skin of the abdomen of a normal person causes a reflex contraction of the abdominal muscles They do not appear until between the 6th and 8th months of age when the infant can sit up unsupported These reflexes are, of course, abolished by any lesion which interrupts afferent conduction from the skin of the abdomen or destroys the anterior horn cells or the motor pathways concerned with the abdominal musculature These reflexes also are often abolished by a lesion which interrupts the corticospinal pathways above the 7th thoracic segment, but they are not permanently abolished by extirpation of the precentral gyrus These confusing facts have thrown se-

is, all types of sensation on the opposite side of the body tend to be lost or grossly affected In addition to the loss or diminution of sensation, lesions of the thalamus are not infrequently associated with spontaneous pain and with marked hypersensitivity which has a very unpleasant quality These manifestations are part of the thalamic syndrome described below and are not well understood

The Thalamic Syndrome of Dejerine and Roussy. This is a characteristic picture occasionally seen in thalamic disease The following are its chief features

1) Asterognosis and sensory ataxia are due to the impairment of cutaneous and kinesthetic senses These and the other effects given below are mainly on the contralateral side of the body

2) There is some loss of tactile and thermal sensations over the body and face, and the threshold for these sensations, and for pain, is frequently raised Sometimes the patient with a thalamic lesion is unable with closed eyes to localize the position of a limb, and must grope in the air in order to find it (*thalamic phantom limb*), or he may have the illusion that the limb is not there at all (*amelognosia*) He also is commonly unable to localize a point which has been stimulated even though he recognizes the stimulus

3) Spontaneous pain occurs in paroxysms and is often excruciating Other stimuli are often associated with an unpleasant sensation Cold in particular may be most disagreeable but light touch may also be unpleasant These spontaneous discomforts and unusual reactions to various stimuli are usually referred to as *thalamic overreaction*

4) Lesions in the thalamus are frequently associated with various abnormal involuntary movements Choreoathetosis and intention tremor are the more common It is believed that choreoathetosis results from the involvement of the projection system from the basal ganglia to the anterior part of the ventrolateral nucleus of the thalamus, whereas an intention tremor arises when the pathway from the dentate nucleus of the cerebellum and the red nucleus is damaged at its point of termination in the posterior part of the ventrolateral nucleus This nucleus in turn projects to the precentral motor cortex and it is thought possible that these abnormal involuntary movements arise because regulating impulses from the thalamus to the motor cortex have been destroyed (Bucy, 1942) Furthermore, such lesions commonly give rise to an abnormal posture of the extended hand known as the "*thalamic hand*" This is characterized by a moderate flexion of the wrist and a hyperextension of the fingers The same posture of the outstretched hand is also referred to by some as the "athetoid hand"

5) Hemiparesis and defects in the visual fields which may be seen in some patients exhibiting the thalamic syndrome are not the result of thalamic involvement but are due to damage to the neighboring internal capsule and geniculocalcarine pathways

DISORDERS OF THE HYPOTHALAMUS

The effects which may result from lesions (e g , tumors, encephalitis, etc) involving the hypothalamic region fall into the following groups *a*) disturbances in regulation of energy exchange or of water balance, *b*) disorders of sleep, drowsiness, somnolence, and less commonly, abnormal wakefulness, or reversal of the sleep rhythm, *c*) emotional manifestations, laughing, crying, or a state resembling "sham rage" in animals, *d*) phenomena attributable to sympathetic or parasympathetic stimulation, and *e*) disorders of the sexual functions and other signs of failure of normal hypothalamic control of the anterior lobe of the pituitary gland

Any one of the foregoing groups of effects may dominate the clinical condition to give rise to one or other of the following syndromes *a*) diabetes insipidus (see p 9–135), *b*) obesity with genital dystrophy (sect 7, chap 2, and *c*) narcolepsy

Narcolepsy. This is the term applied to a disturbance in the sleep mechanism in which sudden attacks of an irresistible desire for sleep occur during the daytime The duration of the attacks, which resemble normal sleep, is quite brief—from a few seconds to 20 min or so The condition may be a sequel to influenza or to epidemic encephalitis involving the hypothalamus or may result from a tumor or injury in this region In other instances the condition appears without known cause (idiopathic narcolepsy) Although evidence is not conclusive, it is very likely that in these latter, also, disordered hypothalamic function is responsible, for other features, e g , obesity, polyuria, or impairment of the sexual functions, pointing to an abnormality of this region are frequently present

Cataplexy[1] is the term given to a condition allied to, and very frequently associated with narcolepsy in which the patient as a result of some emotion—amusement, anger, fear, embarrassment, or surprise—is seized with complete muscular relaxation and weakness The attack is brief, leasing for a few seconds, or for a minute or two at the most Consciousness is not lost but the muscles are completely toneless and powerless for the time, and if the attack supervenes while the subject is standing his knees fail him and he sinks to the ground The deep reflexes are lost Mirth is especially likely to precipitate a cataplectic attack Though narcolepsy occurs without cataplexy the converse is extremely rare This association of the two conditions at once suggests a common pathogenesis, but the muscular atonicity characterizing the cataplectic attack cannot be explained upon any physiological basis

MANIFESTATIONS OF CEREBELLAR DISEASE

The signs of cerebellar disease are much more pronounced in acute lesions, e g , abscess, hemorrhage, or trauma, then in those, e g , tumor, which develop more gradually In slowly developing lesions or after the subsidence of acute cerebellar disease a certain degree of compensation for the cerebellar defect always occurs The majority of the signs of cerebellar disease are the result fundamentally of asynergia of the voluntary muscles

The more conspicuous signs of a cerebellar lesion are *1) Hypotonia of the skeletal muscles 2) Asthenia*, easy fatigability and slowness of movement *3) Dysmetria*, which is the inability to adjust the strength of a contraction required to accomplish a given act For example, when the subject is asked to touch a point with his finger he overshoots the mark, or, less usually, fails to reach it *4) Rebound phenomenon* If the patient is asked to attempt a movement against a resistance which is then suddenly removed, the limb moves forcibly in the direction toward which the effort was made For example, if the observer holds the wrist of the patient, then asks him to bend the elbow, and while he is making the effort, the forearm is suddenly released, flexion of the arm occurs with an unusual degree of force This rebound phenomenon can be attributed to the absence of the "braking" action of the antagonistic muscles 5) *Adiadochokinesis* (a = privative, diadocho = succession, kinesis = movement) is the name given by Babinski to the inability to execute alternating movements rapidly, e g , pronation and supination of the forearm, or flexion and extension of the fingers *6) Deviation of the eyes, and nystagmus 7) Gait* may be staggering, reeling or lurching, often with ataxia of the limbs The *attitude* of the body may be abnormal, the trunk being flexed toward the side of the lesion and the head rotated with the chin pointing toward the sound side *8) Disturbances of speech*, which may be drawling, scanning, sing-song, or explosive in character *9) Intention tremor* that makes it impossible for the patient to perform normal voluntary movements Most of these symptoms have been considered in chapter 5 in connection with normal cerebellar function

DISEASES OF THE BASAL GANGLIA

As noted in chapter 5, the chief clinical features of disease of the basal ganglia are *a)* muscular rigidity resulting in disturbances of posture and movement, *b)* involuntary movements, e g , tremor, athetosis, chorea, *c)* a loss of automatic and associated movements such as those of emotional expression, the swinging of the arms of walking, etc , and *d)* absence of a paralysis of voluntary movement, although it may be distinctly impaired by one of all of the other manifestations The basal ganglia are particularly susceptible to certain toxins (e g , carbon disulfide), to anoxia, and to certain encephalitic processes The following are some of the syndromes met with *a)* Progressive hepatolenticular degeneration, *b)* parkinsonism —paralysis agitans, *c)* chorea, *d)* athetosis

Progressive hepatolenticular degeneration (Wilson's disease). This was described by Wilson in 1912 It is a progressive disease that is familial, characterized by a complex deficiency in copper transport in the blood The symptoms are caused by deposition of copper in the lentiform nucleus The following are its chief features

1) Muscular rigidity is widespread and progressive, it involves face, trunk, and limbs Flexors as well as extensors are affected, but the former more conspicuously than the latter The hypertonus offers a "lead-pipelike" resist-

[1] This should not be confused with catalepsy, an entirely different condition

ance to passive movement and results in slowness and difficulty of movement Eventually, contractures occur and the patient is rendered almost immobile The rigidity of the facial muscles gives a fixed, blank expression The mouth is sometimes held widely open, the smile or laugh is peculiarly stiff and vacuous The hypertonus of the muscles of articulation and deglutition leads to dysarthria (p 9–160) and dysphagia The rigidity is temporarily abolished by the injection of novocaine into the muscles

2) *Involuntary movements* These consist chiefly of tremor (about 6 oscillations/sec) which is increased by excitement or any attempt at voluntary movement, sometimes athetoid movements occur (see below)

3) *Reflexes are normal* There are no sensory changes and although the muscles often show some weakness and are easily fatigued there is no actual paralysis

4) *Cirrhosis (multilobular) of the liver* is found at autopsy, but during life there may be no signs of liver disease In some instances, however, symptoms pointing to the liver precede the nervous manifestations

5) *Emotionalism* Involuntary laughing or crying, and some mental deterioration

6) *Greenish brown pigmentation of the cornea* (in Descemet's membrane) occurs in most cases

Degeneration of the cells of the putamen and globus pallidus, sometimes with cavitation, is found at autopsy The caudate nucleus is affected to a much less degree

The Parkinsonian Syndrome. The principal features of this syndrome are the following *a)* A *coarse tremor* involving head and limbs The hand may show "pill-rolling" movements, i e , rhythmical movements of thumb upon the first two fingers Alternating movements of flexion and extension at the wrist, or of supination and pronation of the forearm, are frequently present When the limb is engaged in some voluntary act the movements often temporarily disappear in that limb The tremor becomes more pronounced during emotional excitement or when the patient is under a nervous strain It disappears during sleep The tremor also disappears on the paralyzed side if hemiplegia supervenes Ablation of the motor area of the cortex, or section of the lateral corticospinal tracts also abolishes it in the related extremities It therefore seems to be dependent upon an intact pyramidal tract However, this has been questioned by some surgeons who believe they have abolished the tremor by producing lesions in the globus pallidus, the thalamus, or the midbrain *b) Muscular rigidity* which leads to slowness and stiffness of movement The rigidity is different both in its quality and distribution from that characteristic of the decerebrate animal or of the hemiplegic patient Flexors and extensor muscles are affected about equally, and the resistance of a limb to passive movements has been described as resembling cogwheels moving slowly upon one another, as though groups of muscles gave way in succession to the stretching force This cogwheel phenomenon is undoubtedly merely a manifestation of the rhythmical innervation of the involved muscles which also gives rise to the tremor The upper limbs are held in characteristic attitudes of adduction at the shoulders, flexion at the elbows, flexion or slight extension at the wrists, flexion at the interphalangeal joints *c)* The *gait* is slow and shuffling with short steps, or it may be "festinating" in character, i e , the patient is bent forward and hastens along with short quick steps as though trying to "catch up to his center of gravity" and prevent his falling When pushed forward or backward he cannot stop quickly but moves by a series of small rapidly repeated steps in the direction in which he is pushed *Propulsion* and *retropulsion* are the respective terms applied to these forward and backward movements There is no true paralysis, the reflexes and sensation are unaffected *d)* One of the most constant manifestations of parkinsonism is the *loss of* certain automatic and *associated movements* As a result the patient does not swing his arms as he walks, the typical facial expressions of various emotions are decreased or absent and the face is said to be masklike Because the many little movements of the body which characterize normal human activity are absent the patient stands rigidly and without moving, "like a statue "

The pathogenesis of these symptoms and their relief following administration of dihydroxyphenylalanine (dopa) have been discussed in chapter 5, pp 9–110–9–112

Chorea. There are two principal forms of this condition

1) Sydenham's Chorea (or St Vitus's dance) is one of the manifestations of rheumatic fever

Its chief feature is involuntary jerky movements, semipurposeful in character, involving the muscles of the limbs and face Facial grimacing, an inability to remain quiet, the frequent dropping of objects from the hands, and the inability to maintain sustained muscular contractions characterize the disease

Sydenham's chorea may be bilateral or unilateral Death is rare and there is consequently uncertainty concerning its neuropathology In those few cases where examinations have been made the lesions have been multiple and diffuse, involving both cerebral cortex and subcortical structures

2) *Huntington's Chorea* is a rare familial disease which results in severe disability and ultimately terminates fatally It is transmitted as a dominant characteristic but the onset is usually late in the fourth decade of life Pathologically one finds severe degeneration and atrophy of the caudate and lenticular nuclei The small cells of the caudate nucleus and putamen are particularly affected However, the large cells are not entirely spared and changes are seen in the globus pallidus and cerebral cortex The disease is characterized by facial grimacing, jerky gesticulating movements of the upper extremities and a lurching, uncertain gait Speech is dysarthric The motor disturbances steadily increase until the patient is incapacitated There is mental deterioration with irritability and occasional homicidal and suicidal tendencies

Athetosis is a form of abnormal involuntary movements beginning most often in childhood It is associated with damage to the basal ganglia, particularly the caudate nucleus and the putamen This damage may result from asphyxia or injury at birth, or from injury or encephalitis subsequently Pure athetosis consists of slow, writhing, twisting movements of the involved extremities and the face In many cases there are also quicker, jerky involuntary movements and the condition is then termed *choreoathetosis* As with most other abnormal involuntary movements these are present during all waking hours and are absent during sleep They are aggravated by nervous tension and are diminished or abolished by the ingestion of alcohol In many patients choreoathetoid movements are unilateral and are associated with a hemiparesis on the same side Unilateral choreoathetosis can be abolished by removal of the contralateral precentral motor cortex (areas 4 and 6) They can also be diminished by section of the central portion of the contralateral cerebral peduncle which contains the corticospinal fibers from the precentral gyrus A few cases have been reported in which the movements have been decreased after surgical lesions have been placed in the region of the internal segment of the globus pallidus or the ansa lenticularis

DISSEMINATED (MULTIPLE) SCLEROSIS

This is a chronic disease of the nervous system characterized anatomically by the occurrence of small patches of demyelination followed by overgrowth of glial tissue throughout the white substance of the brain (especially in the regions beneath the lateral ventricles), and spinal cord The lesions in the cord are most numerous in the corticospinal tracts and in the posterior columns Clinically, the disease is marked by an insidious onset, irregular course, with remissions and relapses of unpredictable duration, and a great variability of signs and symptoms between individual cases Progressive spastic weakness of the muscles of the legs, with increased tendon jerks, "extensor" plantar response (Babinski) and loss of abdominal reflexes, is a common manifestation of this disease Many patients also complain of paresthesias and have a loss of vibration and muscle senses from lesions in the posterior columns The loss of tactile and pain and thermal sensibilities is seldom striking and rarely of an enduring nature Impairment of vision, or even blindness, may result from involvement of the optic nerve, chiasma, or tract Nystagmus, slurring of speech, and intention tremor, once regarded as diagnostic of the disease, appear late if at all and are manifestations of involvement of cerebellar pathways

Epilepsy

Epilepsy is a disorder arising as the result of an abnormal and severe discharge of nerve impulses from some part of the central nervous system, usually the cerebral cortex

TYPES OF SEIZURES

Epilepsy may be generalized, involving the entire organism, or localized Generalized seizures are of three types, grand mal, petit mal, and psychomotor

In *grand mal* there is an abrupt loss of consciousness and a generalized convulsion Two stages of the attack or seizure are recognized In the first or *tonic stage* the muscles contract tonically, the spasms often twisting the facial features and holding the head and limbs in distorted positions The arms are most commonly flexed and the lower limbs rigidly extended After a few seconds the tonic spasm gives place to jerking movements, often violent, of the limbs, face, and muscles of mastication This is spoken of as the *clonic stage* Either during this stage or in the tonic stage the tongue may be bitten Before the onset of the convulsion a large proportion of epileptics receive a warning in the form of a sensation or hallucination, the character of which varies in individual cases The warning sensation or *aura*, as it is called, may be auditory (voices, music) visual (flashes of light, sparks) olfactory, gustatory, cutaneous, visceral or vasomotor, equilibratory, or anesthetic, i e , a numbness in some part of the body Turning of the head and trunk to one side and deviation of the eyes are commonly observed The patient sometimes utters a cry or scream—the epileptic cry—just before consciousness is lost After the convulsion the subject remains for a time in a stupor Sometimes a number of convulsive seizures occur in rapid succession, the patient failing to regain consciousness in the intervals between them This very serious condition is called *status epilepticus*

Minor seizures known as *petit mal* are characterized by brief periods during which the patient loses contact with his environment He stares into space There are no convulsive movements

A third form known as *psychomotor* epilepsy is marked by automatic movements, such as smacking of the lips, chewing, together with a clouded, "dreamy" feeling of unreality, or of having seen or heard before some sight or auditory impression which is actually happening at the moment (*déjà vu* or *déja entendu* phenomenon, respectively) Or there may be a confused mental state persisting for a minute or two, or for a longer period. During this time the patient may perform automatic acts of which he is quite unaware and does not remember Generalized convulsions do not occur nor does the subject fall to the ground during the psychomotor seizure, but patients who have such seizures may also suffer from generalized convulsions Psychomotor seizures are most commonly due to disturbances in the temporal lobe

Localized epilepsy may occur in several different forms, these are often referred to as *jacksonian epilepsy* They signify an abnormality in a particular part of the brain Thus, motor Jacksonian seizures are indicative of a disturbance in the precentral motor cortex, hallucinations of taste and smell are indicative of an abnormal nervous discharge from the uncus on the medial surface of the temporal lobe, sensory jacksonian seizures indicate disease in the parietal region, etc

Motor jacksonian seizures commonly begin by a turning of the head and eyes toward the side opposite the cortical lesion if it is located in the anterior part of the precentral motor cortex, or by a twitching of the foot or leg if it is in the upper part of the precentral region Such attacks may also begin with a twitching of a finger, or of the hand or the face depending upon the seat of the irritation in the cortex which initiates the attack These convulsions may remain localized to the part of the body in which they start or they may spread by a steady "march" to neighboring parts as portions of the cortex lying near to the original focus become involved in the irritative process The spreading convulsion may cease after extending to all of one extremity or to one side of the body or it may continue until a generalized convulsion has occurred

Such focal or localized convulsive seizures or seizures with localized onsets may develop as the result of many different kinds of disease processes in the brain Tumors, abscesses, traumatic scars, vascular abnormalities are among the more common processes found to be responsible for such attacks It must be noted that not all patients who have such disease processes suffer from epileptic attacks and that even in those patients who do have them the attacks occur at relatively infrequent intervals although the disease process is constantly present It is thus obvious that some other process, neurological or metabolic, is essential to the development of such seizures The tumor or scar alone is not enough

Idiopathic or cryptogenic epilepsy. Epilepsy which cannot be explained by the presence of a demonstrable lesion of the brain is called idiopathic or cryptogenic Though many theories have been advanced, the cause of this type of

the disease remains obscure There is a tendency today to look upon epilepsy as a symptom, or rather a group of symptoms common to several rather than to a single primary pathological state Convulsions, whose features are indistinguishable from those of the epileptic seizure, occur in a number of conditions In animals convulsions may be produced by injections of absinthe or caffeine Hypocalcemia, hypoglycemia, cerebral edema, or anemia and other states may be accompanied by generalized convulsions of an epileptiform character

ORIGIN OF SEIZURES

With regard to the neural mechanism through which the fits are produced, some authorities have thought that all convulsions are the result of increased excitability of the cortex and, therefore, comparable in their mode of production to those of the Jacksonian type, or to those produced by experimental stimulation of the cortex Others view the convulsions as a release phenomenon due to the inhibition of cortical areas which normally exert a controlling influence upon lower motor centers, e g , weakening of suppressor action Finally, there is the view of Penfield and Jasper (1954) that the primary disturbance in idiopathic epilepsy has its origin in a subcortical mechanism

The close correspondence between the convulsive seizure as seen in focal epilepsy or after artificial stimulation of the cortex, and the fits of idiopathic epilepsy appears to support the first alternative as a cause of the generalized seizure known as grand mal However, the *primary* disturbance responsible for the cortical discharge in idiopathic epilepsy probably lies outside the cortex Evidence for a subcortical origin of petit mal attacks is provided by electroencephalographic studies During such an attack, electrical potentials of the wave-spike type appear synchronously over prefrontal areas as well as from other widely separated cortical areas This suggests that the cortical discharges are initiated in a common subcortical region (centrencephalic system) connected with extensive areas of the cortex

The nature of the underlying bodily state responsible for the convulsive seizures of idiopathic epilepsy are unknown There are a number of drugs, e g , metrazol, which are capable of inducing convulsions, and several others, e g , phenobarbital and tridione, which are anticonvulsive Susceptibility to an epileptic seizure is influenced by several physiological and biochemical factors which are given in Table 9 7, but they do not bear a specific relationship to the cause of idiopathic epilepsy, but merely affect the tendency to convulsive seizures of any kind

Table 9 7
Summary of conditions which tend to predispose to or to prevent epileptic seizures

Condition	Prevents Seizures	Precipitates Seizures
Oxygen	Rich supply	Poor supply
Acid-base equilibrium	Acidosis Ingestion of acids or acid-forming salts Breathing CO_2	Alkalosis Hyperpnea—"blowing off" CO_2
Water balance	Dehydration	Edema
Serum calcium	Increase	Decrease
Blood glucose	Normal level	Decrease (hypoglycemia)
Body temperature		Increase

The view that the actual seizure was initiated by cerebral ischemia resulting from a spasm of the cortical vessels has not been sustained The blood flow through the brain of epileptics and the oxygen consumption have been found to be no different from those of nonepileptics In focal epilepsy, however, local ischemia may play a part Nevertheless, biochemical studies of such epileptogenic areas have not revealed any important abnormality, and the pH of the extracellular fluid shows no change from the normal, or a slight rise

The induction of a dehydrated state of acidosis (by means of a ketogenic diet) has been employed to reduce the susceptibility of epileptics to seizures but not very effectively, and pitressin combined with an increase in the water intake is used as a test (*water-pitressin test*) in diagnosis But in only about 30% of epileptics is a seizure precipitated, so a negative result does not exclude epilepsy

BIBLIOGRAPHY

ADOLPH E F Early concepts of physiological regulations *Physiol Rev* 41 737-770 1961

ADRIAN, E D *The Mechanism of Nervous Action* Philadelphia Univ Pennsylvania Press 103 pp 1935

ADRIAN F D Double representation of the feet in the sensory cortex of the cat *J Physiol (London)* 98 16P-18P 1940

ADRIAN E D AND D W BRONK The discharge of impulses in motor nerve fibres Part II The frequency of discharge in reflex and voluntary contractions *J Physiol (London)* 67 119-151 1929

ADRIAN, E C AND B H C MATTHEWS The Berger rhythm Potential changes from the occipital lobes in man *Brain* 57 355-385 1934

AGRANOFF B W Recent studies on the stages of memory formation in the goldfish In *Molecular Approaches to Learning and Memory* edited by W L Byrne New York Academic Press p 35-39 1970

AHLQUIST R P A study of the adrenotropic receptors *Amer J Physiol* 153 586-600 1948

AKERT K W P KOELLA AND R HESS Sleep produced by electrical stimulation of the thalamus *Amer J Physiol* 168 260 267 1952

ALBE FESSARD D C ROCHA MIRANDA, AND F OSWALDO CRUZ Activites d origine somesthesique evoquees au niveau du cortex non specifique et du centre median du thalamus chez le singe anesthesie au chloralose *Electroenceph Clin Neurophysiol* 11 777 787 1959

ALBE FESSARD D AND A ROUGEUL Activites d origine somesthesique évoquées sur le cortex nonspecifique du chat anesthésie au chloralose role du centre median du thalamus *Electroenceph Clin Neurophysiol* 10 131 152 1958

AMASSIAN V Studies on organization of a somesthetic association area including a single unit analysis *J Neurophysiol* 17 39 58 1954

ANAND, B K, AND J R BROBECK Hypothalamic control of food intake in rats and cats *Yale J Biol Med* 24 123 140 1951

ANONYMOUS Probing the brain *Newsweek* 72 60 62, 1971

AREY, L B *Developmental Anatomy*, Ed 7 Philadelphia Saunders, 695 pp 1965

ARIENS KAPPERS, C U *The Evolution of the Nervous System in Invertebrates Vertebrates and Man* Haarlem, Netherlands Bohn 335 pp 1929

AXELROD J, AND I J KOPIN The uptake storage, release and metabolism of noradrenaline in sympathetic nerves. *Progr Brain Res* 31 21 32 1969

AXELSSON J AND S THESLEFF A study of supersensitivity in denervated mammalian skeletal muscle *J Physiol (London)* 149 178 193, 1957

BABICH, F R, A I JACOBSON, S BUBASH, AND A JACOBSON Transfer of a response to naive rats by injection of ribonucleic acid extracted from trained rats *Science* 149 656 657 1965

BABINSKI, M J Sur le reflexe cutané plantaire dans certaines affections organiques du systeme nerveux central *C R Soc Biol (Paris)* 48 207-208, 1896

BAILEY, P, AND G VON BONIN *The Isocortex of Man* Urbana, Ill Univ Illinois Press, 301 pp, 1951

BARBOUR, H G Die Wirkung unmittelbarer Erwarmung und Abkuhlung der Warmezentra auf die Korpertemperatur *Arch Exp Pathol Pharmakol* 70 1-26, 1912

BARD, P A diencephalic mechanism for the expression of rage with special reference to the sympathetic nervous system *Amer J Physiol* 84 490-515, 1928

BARD, P, AND C McC BROOKS Localized cortical control of some postural reactions in the cat and rat together with evidence that small cortical remnants may function normally *Res Publ Ass Res Nerv Ment Dis* 13 107-157 1934

BARD, P, AND M B MACHT The behavior of chronically decerebrate cats In *Ciba Foundation Symposium on the Neurological Basis of Behavior*, edited by G E W Wolstenholme and C M O Connor Boston Little, Brown p 55–75, 1958

BARD, P., AND V B MOUNTCASTLE Some forebrain mechanisms involved in expression of rage with special reference to suppression of angry behavior *Res Publ Ass Res Nerv Ment Dis* 27 362-404 1948

BARKER, D (ed) *Symposium on Muscle Receptors* Hong Kong Hong Kong Univ Press 293 pp, 1962

BARONDES, S H Some critical variables in studies of the effect of inhibitors of protein synthesis on memory In *Molecular Approaches to Learning and Memory* edited by W L Byrne New York Academic Press p 27-34 1970

BARRON D H AND B H C MATTHEWS The interpretation of potential changes in the spinal cord *J Physiol (London)* 92 276-321 1938

BEATON L E C LEININGER W A McKINLEY H W MAGOUN AND S W RANSON Neurogenic hyperthermia and its treatment with soluble pentobarbital in the monkey *Amer Med Ass Arch Neurol Psychiat* 49 518-536 1943

BEKHTEREV V M *General Principles of Human Reflexology* translated from Russian by Emma and W Murphy London Jarrolds 467 pp 1933

BELL C *Idea of a New Anatomy of the Brain Submitted for the Observations of His Friends* London Privately printed 36 pp 1811

BENZINGER T H Heat regulation Homeostasis of central temperature in man *Physiol Rev* 39 671-759 1969

BERGER, H Über das Elektrenkephalogram des Menschen *Arch Psychiat Nervenkr* 87 527-570 1929

BERMAN A L Interaction of cortical responses to somatic and auditory stimuli in anterior ectosylvian gyrus of cat *J Neurophysiol* 24 608 620 1961

BERNARD C *Lecons sur les Phenomenes de la Vie Communs aux Animaux et aux Vegetaux* Paris Baillhere p 112-379 1878

BLACK A H, N J CARLSON, AND R L SOLOMON Exploratory studies of the conditioning of autonomic responses in curarized dogs *Psychol Monogr* 76 no 29 (whole no 548), 31 pp 1962

BLACK, P (ed) *Physiological Correlates of Emotion* New York Academic Press, 309 pp 1970

BLOMFIELD S, AND D MARR How the cerebellum may be used *Nature (London)* 227 1224-1228 1970

BLOOM, W, AND D W FAWCETT, *A Textbook of Histology* Philadelphia W B Saunders 858 pp, 1968

BODIAN, D Neurons, circuits, and neuroglia In *The Neurosciences A Study Program* edited by G C Quarton, T Melnechuk, and F O Schmitt New York Rockefeller Univ Press p 6-24, 1967

BOYD I A The structure and innervation of the nuclear bag muscle fibre system and the nuclear chain muscle fibre system in mammalian muscle spindles *Phil Trans Roy Soc B* 245 81-136 1962

BREMER, F Cerveau "isolé" et physiologie du sommeil *C R Soc Biol (Paris)* 122 464-467, 1936

BRICKNER R M An interpretation of frontal lobe function based upon the study of a case of partial bilateral frontal lobectomy *Res Publ Ass Res Nerv Ment Dis* 13 259-351 1934

BROCK, L G, J S COOMBS, AND J C ECCLES The recording of potentials from motoneurones with an intracellular electrode *J Physiol (London)* 117 431-460, 1952

BRODMANN, K *Vergleichende Lokalisationlehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues* Leipzig Barth 324 pp, 1909

BUCY, P A Neural mechanisms in athetosis and tremor *J Neuropath & Exp Neurol* 1 224-239 1942

BUCY, P C Is there a pyramidal tract? *Brain* 80 376-392, 1957

BUCY P C, AND J F FULTON Ipsilateral representation in the motor and premotor cortex of monkeys *Brain* 56 318-342, 1933

BULLOCK, T H Signals and neuronal coding In *The Neurosciences A Study Program*, edited by G C Quarton, T Melnechuk, and F O Schmitt New York Rockefeller Univ Press, p 347-352, 1967

BULLOCK, T H, AND G A HORRIDGE Structure and function in the nervous systems of invertebrates San Francisco Freeman, vol I, p 1-798, vol II, p 799-1719, 1965

BUSER, P, P BORENSTEIN, AND J BRUNER Étude des systèmes "associatifs" visuels et auditifs chez le chat anesthésie au chloralose *Electroenceph Clin Neurophysiol* 11 305-324, 1959

BYRNE, W L *Molecular Approaches to Learning and Memory* New York Academic Press, 358 pp 1970

CAMPBELL A W *Histological Studies on the Localization of Cerebral Function* New York Cambridge Univ Press 360 pp 1905

CANNON, W B The physiological basis of thirst *Proc Roy Soc (London) [Biol] Ser B* 90 283-301, 1918

CANNON, W B *Bodily Changes in Pain, Hunger Fear and Rage* New York Appleton 311 pp 1929

CANNON, W B Organization for physiological homeostasis *Physiol Rev* 9 399-431 1929

CANNON, W B *The Wisdom of the Body* New York Norton 312 pp, 1932

CANNON W B AND Z M BACQ Studies on the conditions of activity in endocrine organs XXVI A hormone produced by sympathetic action on smooth muscle *Amer J Physiol* 96 392-412 1931

CANNON, W B AND S W BRITTON Studies on the conditions of activity in the endocrine glands XV Pseudaffective medulliadrenal secretion *Amer J Physiol* 72 283-294, 1925

CANNON, W B H F NEWTON, E M BRIGHT V MENKIN, AND R M MOORE Some aspects of the physiology of animals surviving complete exclusion of sympathetic nerve impulses *Amer J Physiol* 89 84-107 1929

CANNON, W B AND A ROSENBLUETH Studies on conditions of activity in endocrine organs XXIX Sympathin E and sympathin I *Amer J Physiol* 104 557-574, 1933

CARLSSON, A, M LINDQUIST, AND T MAGNUSSON 3 4 dihydroxyphenylalanine and 5 hydroxytryptophan as reserpine antagonists *Nature (London)* 180 1200, 1957

CARLSSON, A M LINDQUIST, T MAGNUSSON AND B WALDECK On the presence of 3 hydroxytyramine in brain *Science* 127 471, 1958

CARPENTER, M B, AND J R WHITTIER Study of methods for producing experimental lesions of the central nervous system with special reference to stereotaxic technique *J Comp Neurol* 97 73-132, 1952

CATON, R Interim report on investigation of the electric currents of the brain In *Experiments on the Biliary Secretion of the Dog* London 100 pp 1875 Also published as *Brit Med J* (Suppl), 1875-1877

CHANG, H-T, T C RUCH, AND A A WARD, JR Topographical representation of muscles in motor cortex of monkeys *J Neurophysiol* 19 39-56, 1947

CLARK, G, H W MAGOUN, AND S W RANSON Hypothalamic regulation of body temperature *J Neurophysiol* 2 61-80, 1939

CLARKE, E, AND C D O'MALLEY *The Human Brain and Spinal Cord* Berkeley Univ California Press, 926 pp, 1968

COHN, D, AND D JOSEPH Influence of body weight and body fat on appetite of "normal" lean and obese rats *Yale J Biol Med* 34 598-607, 1962

COPENHAVER, W M, R P BUNGE, AND M B BUNGE *Bailey's Textbook of Histology*, Ed 16 Baltimore Williams & Wilkins, 745 pp, 1971

COTZIAS, G C Metabolic modification of some neurologic disorders *J A M A*, 210 1255-1262, 1969

COTZIAS, G C, P S PAPAVASILIOU, AND R GELENE Modification of Parkinsonism—chronic treatment with L-DOPA *New Eng J Med* 280 337-345, 1969

COTZIAS, G C, M H VAN WOERT, AND L M SCHIFFER Aromatic amino acids and modification of Parkinsonism *New Eng J Med* 276 374-379, 1967

CREED, R S, D DENNY-BROWN, J C ECCLES, E G T LIDDELL, AND C S SHERRINGTON *Reflex Activity of the Spinal Cord* New York Oxford Univ Press, 183 pp, 1932

CROSBIE, R J, J D HARDY, AND E FESSENDEN Electrical analog simulation of temperature regulation in man IRE Trans Bio Med Elect 8 245-252, 1961

CROSS, B A, AND J D GREEN Activity of single neurones in the hypothalamus Effect of osmotic and other stimuli *J Physiol (London)* 148 554-569 1959

CURTIS, D R, A W DUGGAN, D FELIX, AND G A R JOHNSTON GABA, bicuculline and central inhibition *Nature (London)* 226 1222-1224, 1970

CUSHING, H A note upon the faradic stimulation of the postcentral gyrus in conscious patients *Brain* 32 44-53, 1909

DALE, H H On some physiological actions of ergot *J Physiol (London)* 34 163-206, 1906

DALE H H Nomenclature of fibres in the autonomic system and their effects *J Physiol (London)* 80 10P-11P 1933

DALE, H H AND W FELDBERG The chemical transmission of secretory impulses to the sweat glands of the cat *J Physiol (London)* 82 121-127 1934

DAVIS, R A Victorian physician scholar and pioneer physiologist *Surg Gynec Obstet* 119 1333-1340, 1964

DE KLEIJN, A Experimental physiology of the labyrinth *J Laryng* 38 646-663, 1923

DEL CASTILLO J AND B KATZ Local activity at a depolarized nerve muscle junction *J Physiol (London)* 128 396-411 1955

DELGADO J M R Telemetry and telestimulation of the brain In *Bio telemetry* edited by Lloyd Slater New York Pergamon Press, p 231-249 1963

DELGADO, J M R Modulation of emotions by cerebral radio stimulation In *Physiological Correlates of Emotion* edited by P Black New York Academic Press chap 9 p 189-202 1970

DENNY BROWN D On the nature of postural reflexes *Proc Roy Soc (London) [Biol] Ser B* 104 252-301 1929

DERBYSHIRE, A J B REMPEL, A FORBES, AND E F LAMBERT The effect of anesthetics on action potentials in the cerebral cortex of the cat *Amer J Physiol* 116 557-596 1936

DE ROBERTIS E D P *Histophysiology of Synapses and Neurosecretion* New York Macmillan, 244 pp, 1964

DE ROBERTIS, E D P, AND R CARREA (eds) *Biology of Neuroglia Progress in Brain Research* Amsterdam Elsevier vol 15, 297 pp, 1965

DETHIER, V G, AND E STELLAR *Animal Behavior*, Ed 3 Englewood Cliffs, N J Prentice Hall, 152 pp, 1970

DINGMAN W, AND M B SPORN The incorporation of 8 aza guanine into rat brain RNA and its effect on maze learning by the rat an inquiry into the biochemical basis of memory *J Psychiat Res* 1 1-11, 1961

DOTY, R W Conditioned reflexes formed and evoked by brain stimulation In *Electrical Stimulation of the Brain* edited by D E Sheer Austin Univ Texas Press, p 397-412, 1961

DREIFUSS, J J, I KALNINS J S KELLY, AND K B RUF Action potentials and release of neurohypophysial hormones *in vitro J Physiol (London)* 215 805-817, 1971

DU BOIS, E F Heat loss from the human body *Bull N Y Acad Med Ser 2*, 15 142-173, 1939

DURUP, G, AND A FESSARD L'électroencéphalogramme de l'homme Données quantitatives sur l'arret provoqué par des stimuli visuels ou auditifs *C R Soc Biol (Paris)* 122 756 758, 1936

DYKMAN, R A, AND P S SHURRAGER Successive and maintained conditioning in spinal carnivores *J Comp Physiol Psychol* 49 27-35, 1956

ECCLES, J C *The Neurophysiological Basis of Mind* Oxford Clarendon Press, 314 pp, 1953

ECCLES, J C *The Physiology of Nerve Cells* Baltimore Johns Hopkins Press, 270 pp, 1957

ECCLES, J C Postsynaptic inhibition in the central nervous system In *The Neurosciences A Study Program*, edited by G C Quarton, T Melnechuk, and F O Schmitt New York Rockefeller Univ Press, p 408-427 1967

ECCLES, J C, D S FABER, J T MURPHY, N H SABAH, AND H TABORIKOVÁ Investigations on integration of mossy fiber inputs to Purkinje cells in the anterior lobe *Exp Brain Res* 13 54-77, 1971

ECCLES, J C, M ITO, AND J SZENTAGOTHAI *The Cerebellum as a Neuronal Machine* New York Springer Verlag, 335 pp, 1967

EISENMAN, J S Pyrogen induced changes in the thermosensitivity of septal and preoptic neurons *Amer J Physiol* 216 330 334, 1969

EPSTEIN, A N, J T FITZSIMONS, AND B J ROLLS Drinking induced by injection of angiotensin into the brain of the rat *J Physiol (London)* 210 457 474, 1970

EPSTEIN, A N, D SPECTOR, A SAMMAN, AND C GOLDBLUM Exaggerated prandial drinking in the rat without salivary glands *Nature (London)* 29 1342-1343, 1964

EPSTEIN, A N, AND P TEITELBAUM Severe and persistent deficits in thirst produced by lateral hypothalamic damage In *Thirst*, edited by M J Wayner Oxford Symposium Publications Division, Pergamon Press, p 395-410, 1964

ERLANGER, J, AND H S GASSER *Electrical Signs of Nervous Ac-*

tivity Philadelphia Univ Pennsylvania Press 221 pp , 1937
EVARTS F V Representation of movements and muscles by pyramidal tract neurons of the precentral motor cortex In *Neurophysiological Basis of Normal and Abnormal Motor Activities* edited by M D Yahr and D P Purpura New York Raven Press p 215 251 1967
FATT, P AND B KATZ Spontaneous subthreshold activity at motor nerve endings *J Physiol (London)* 117 109 128 1952
FISHER C W R INGRAM AND S W RANSON Diabetes insipidus and the neuro hormonal control of water balance A contribution to the structure and function of the hypothalamico hypophyseal system Ann Arbor Mich Edwards Brothers 212 pp 1938
FITZSIMONS J T The role of a renal thirst factor in drinking induced by extracellular stimuli *J Physiol (London)* 201 349–368 1969
FLEXNER J B L B FLEXNER AND F STELLAR Memory in mice as affected by intracerebral puromycin *Science* 141 57–59 1963
FOERSTER O The dermatomes in man *Brain* 56 1–39 1933
FOLKOW F AND F H RUBINSTEIN Behavioural and autonomic patterns evoked by stimulation of the lateral hypothalamic area in the cat *Acta Physiol Scand* 65 292 299 1965
FRENCH J D R HERNÁNDEZ PEON AND R B LIVINGSTON Projections from cortex to cephalic brain stem (reticular formation) in monkey *J Neurophysiol* 18 74 95 1955
FULTON J F *Physiology of the Nervous System*, Ed 3 New York Oxford Univ Press 667 pp 1949
FULTON, J F *Frontal Lobotomy and Affective Behavior* New York Norton 159 pp 1951
FULTON, J F AND L G WILSON *Selected Readings in the History of Physiology*, Ed 2 Springfield Ill Thomas, 492 pp 1966
FURCHGOTT R F Dibenamine blockade in strips of rabbit aorta and its use in differentiating receptors *J Pharmacol Exp Ther* 111 265 284, 1954
FUSCO M M J D HARDY, AND H T HAMMEL Interaction of central and peripheral factors in physiological temperature regulation *Amer J Physiol* 200 572 580 1961
GALAMBOS R Suppression of auditory nerve activity by stimulation of efferent fibers to cochlea *J Neurophysiol* 19 424–437 1956
GALAMBOS R, AND G C SHEATZ An electroencephalographic study of classical conditioning *Amer J Physiol* 203 173 184, 1962
GARDNER M R, AND W R ASHBY Connectance of large dynamic (cybernetic) systems Critical values for stability *Nature (London)* 228 784 1970
GASKELL, W H *The Involuntary Nervous System* London Longmans, Green, 178 pp 1916
GAZZANIGA M S *The Bisected Brain* New York Appleton Century Crofts (Meredith), 172 pp , 1970
GEFFEN, L B , AND G B LIVETT Synaptic vesicles in sympathetic neurons *Physiol Rev* 51 98 157, 1971
GELLHORN, E Further investigations on the recovery of inhibited conditioned reactions *Proc Soc Exp Biol Med* 59 155-161, 1945
GEORGE, F H Brain models a brief review *Int J Systems Sci* 1 75 88 1970
GESCHWIND, N The organization of language and the brain *Science* 170 940–944, 1970
GLEES P , AND J COLE Recovery of skilled motor functions after small repeated lesions of motor cortex in macaque *J Neurophysiol* 13 137 148, 1950
GOLTZ, F , AND J R EWALD Der Hund mit verkurztem Rückenmark *Pflueger Arch* 63 362-400 1896
GRANIT, R *Receptors and Sensory Perception* New Haven Yale Univ Press, 369 pp , 1955
GRANIT, R *The Basis of Motor Control* London Academic Press, 346 pp , 1970
GREEN, J D , C A CLEMENTE, AND J DE GROOT Rhinencephalic lesions and behavior in cats *J Comp Neurol* 108 505-545, 1957
GRODINS, F S *Control Theory and Biological Systems* New York Columbia Univ Press, 205 pp , 1963
GRUNDFEST, H Synaptic and ephaptic transmission In *The Neurosciences, a Study Program*, edited by G C Quarton, T Melnechuk, and F O Schmitt New York Rockefeller Univ Press, p 353–372, 1967
HAASE J , AND B VOGEL Direkte und indirekte Wirkungen supraspinaler Reizunger auf Renshaw Zellen *Pfuegers Arch* 325 334–346 1971
HAINSWORTH F R Saliva spreading activity and body temperature regulation in the rat *Amer J Physiol* 212 1288–1292 1967
HALSTEAD W C Brain and intelligence a quantitative study of the frontal lobes Chicago Univ Chicago Press, 206 pp 1947
HAMMEL H T Neurons and temperature regulation In *Physiological Controls and Regulations* edited by W S Yamamoto and J R Brobeck Philadelphia Saunders chap 5 p 71-97 1965
HANBERRY J AND H JASPER Independence of diffuse thalamocortical projection system shown by specific nuclear destruction *J Neurophysiol* 16 252-271 1953
HARDY, J D Heat transfer In *Physiology of Heat Regulation and the Science of Clothing* edited by L H Newburgh Philadelphia Saunders p 78–108 1949
HARDY J D The ' set point concept in physiological temperature regulation In *Physiological Controls and Regulations* edited by W S Yamamoto and J R Brobeck Philadelphia Saunders chap 6 p 98–116 1965
HARE, K Visceral functions of the nervous system *Ann Rev Physiol* 8 375–411 1946
HARMON P J AND M B CARPENTER Volumetric comparisons of the basal ganglia of various primates including man *J Comp Neurol* 93 125–137, 1950
HARMON, I D AND E R LEWIS Neural modeling *Physiol Rev* 46 513–591 1966
HARRIS, G W *Neural Control of the Pituitary Gland* London Edward Arnold, 298 pp 1955
HEAD, H On disturbances of sensation with especial reference to the pain of visceral disease *Brain* 16 1–133, 1893
HEAD H *Aphasia and Kindred Disorders of Speech* Cambridge, Mass Cambridge Univ Press vol 2, 407 pp , 1926
HENRY, J P , O H GAUER, AND J L REEVES Evidence of the atrial location of receptors influencing urine flow *Circ Res* 4 85 90, 1956
HERIOT, J T , AND P D COLEMAN The effect of electroconvulsive shock on retention of a modified ' one trial" conditioned avoidance *J Comp Physiol Psychol* 55 1082–1084, 1962
HERRICK C J The functions of the olfactory parts of the cerebral cortex *Proc Nat Acad Sci (U S A)* 19 7–14, 1933
HESS, R JR W P KOELLA, AND K AKERT Cortical and subcortical recordings in natural and artificially induced sleep in cats *Electroenceph Clin Neurophysiol* 5 75–90, 1953
HESS W R *Diencephalon Autonomic and Extrapyramidal Functions* New York Grune & Stratton, 79 pp 1954
HETHERINGTON A W AND S W RANSON The spontaneous activity and food intake of rats with hypothalamic lesions *Amer J Physiol* 136 609 617, 1942
HILGARD, E R , AND D G MARQUIS *Conditioning and Learning* New York Appleton Century Crofts, 429 pp , 1940
HOBSON, J A Sleep physiologic aspects *New Eng J Med* 281 1343–1345 1969
HOLLINSHEAD, W H *Textbook of Anatomy*, Ed 2 New York Hoeber Harper & Row, 994 pp , 1967
HOLMES G , AND W P MAY On the exact origin of the pyramidal tracts in man and other mammals *Brain* 32 1–54, 1909
HORNYKIEWICZ, O Die topische Lokalisation und das Verhalten von Noradrenalin und Dopamin (3 Hydroxytyramin) in der *substantia nigra* des normalen and Parkinsonkranken Menschen *Wien Klin Wschr* 75 309–312, 1963
HORTON, E W Hypotheses on physiological roles of prostaglandins *Physiol Rev* 39 122–161, 1969
HUXLEY, A F , AND R W STRAUB Local activation and interfibrillar structures in striated muscle *J Physiol (London)* 143 40P-41P, 1958
HUXLEY, A F , AND R E TAYLOR Local activation of striated muscle fibers *J Physiol (London)* 144 426-441, 1958
HYDEN, H A molecular basis of neuron glia interaction In *Macromolecular Specificity and Biological Memory*, edited by F O Schmitt Cambridge, Mass M I T Press, p 55–69, 1962
INGRAM, W R The hypothalamus A review of the experimental data *Psychosom Med* 1 48–91, 1939
INGRAM, W R Brain stem mechanisms in behavior *Electroenceph Clin Neurophysiol* 4 397–406, 1952

IVERSEN L L The catecholamines *Nature (London)* 214 8-14 1967

JACKSON, J H The Lumleian lectures on convulsive seizures *Brit Med J* I 703-707, 765-771 821-827, 1890

JARVIK, M E The role of consolidation in memory In *Molecular Approaches to Learning and Memory* edited by W L Byrne New York Academic Press p 15-26, 1970

JASPER, H H, C AJMONE MARSAN AND J STOLL Corticofugal projections to the brain stem *Arch Neurol Psychiat* 67 155-171 1952

JASPER H H G RICCI AND B DOANE Microelectrode analysis of cortical discharge during avoidance conditioning in the monkey *Electroenceph Clin Neurophysiol* (suppl 13) pp 137-155 1960

JASPER H H AND C SHAGASS Conditioning the occipital alpha rhythm in man *J Exp Psychol* 28 373-388 1941

JOHN E R *Mechanisms of Memory* New York Academic Press 468 pp 1967

JOHN E R Summary Symposium on memory transfer A A A S New York Dec 1967 In *Molecular Approaches to Learning and Memory* edited by W L Byrne New York Academic Press p 335-342 1970

JOUVET, M Neurophysiology of the states of sleep *Physiol Rev* 47 117-177, 1967

KAADA B R, K H PRIBRAM, AND J A EPSTEIN Respiratory and vascular responses in monkeys from temporal pole, insula or bital surface and cingulate gyrus *J Neurophysiol* 12 347-356 1949

KAMIKAWA K, J T McILWAIN AND W R ADEY Response patterns of thalamic neurons during classical conditioning, *Electroenceph Clin Neurophysiol* 17 485-496 1964

KANDEL E R AND W A SPENCER Cellular neurophysiological approaches in the study of learning *Physiol Rev* 48 65-134 1968

KARPLUS J P AND A KREIDL Gehirn und sympathicus II Ein Sympathicuszentrum im Zwischenhirn *Pflueger Arch* 135 401-416, 1910

KATZ B Quantal mechanism of neural transmitter release *Science* 173 123-126, 1971

KENNEDY, D, A I SELVERSTON, AND M P REMLER Analysis of restricted neural networks *Science* 164 1488-1496 1969

KIMBLE, G A *Hilgard and Marquis' Conditioning and Learning* Ed 2 New York Appleton Century Crofts, 590 pp 1961

KLEIBER, M *The Fire of Life* New York Wiley 454 pp, 1961

KLUVER, H AND P C BUCY Preliminary analysis of functions of the temporal lobes in monkeys *Arch Neurol Psychiat (Chicago)* 42 979-1000, 1939

KOIZUMI, K, T ISHIKAWA, AND C McC BROOKS Control of activity of neurons in the supraoptic nucleus *J Neurophysiol* 27 878-892, 1964

KORNER P I Integrative neural cardiovascular control *Physiol Rev* 51 312-367 1971

KRAVITZ, E A Acetylcholine γ aminobutyric acid, and glutamic acid Physiological and chemical studies related to their roles as neurotransmitter agents In *The Neurosciences A Study Program*, edited by G C Quarton T Melnechuk, and F O Schmitt New York Rockefeller Univ Press, p 433-444 1967

LANDAU, W M Evoked potentials In *The Neurosciences A Study Program*, edited by G C Quarton, T Melnechuk and F O Schmitt New York Rockefeller Univ Press, p 469-482, 1967

LANGLEY, J N The autonomic nervous system *Brain* 26 1-26 1903

LANGLEY, J N *The Autonomic Nervous System* Cambridge Hefer, 80 pp 1921

LASHLEY, K S (1929) *Brain Mechanisms and Intelligence* New York Republished by Dover Publications, 200 pp, 1963

LEGENDRE R, AND H C R PIÉRON Le problem des facteurs du sommeil Resultats d'injections vasculaires et intra cerebrales de liquides insomniques *C R Soc Biol (Paris)* 68 1077-1079, 1910

LEIBOWITZ, S F Hypothalamic alpha- and beta adrenergic systems regulate both thirst and hunger in the rat *Proc Nat Acad Sci USA* 68 332-334, 1971

LENNEBERG, E H Brain correlates of language In *The Neurosciences Second Study Program*, edited by F O Schmitt New York Rockefeller Univ Press p 361-371 1970

LINDSLEY, D B, L H SCHREINER W B KNOWLES, AND H W MAGOUN Behavioral and EEG changes following chronic brain stem lesions in the cat *Electroenceph Clin Neurophysiol* 2 483-498 1950

LING G AND R W GERARD The normal membrane potential of frog sartorius fibers *J Cell Physiol* 34 383-396 1949

LLOYD D P C Reflex action in relation to pattern and peripheral source of afferent stimulation *J Neurophysiol* 6 111-119 1943a

LLOYD D P C Conduction and synaptic transmission of reflex response to stretch in spinal cats *J Neurophysiol* 6 317-326, 1943b

LLOYD, D P C Functional organization of the spinal cord *Physiol Rev* 24 1-17, 1944

LLOYD D P C Facilitation and inhibition of spinal moroneurons *J Neurophysiol* 9 421-438, 1946

LOEWI O Uber humorale Ubertragbarkeit der Herznervenwirkung *Pflueger Arch* 189 239-242 1921

LORENTE DE NÓ R Limits of variation of the synaptic delay of motoneurons *J Neurophysiol* 1 187-194 1938a

LORENTE DE NÓ R Synaptic stimulation of motoneurons as a local process *J Neurophysiol* 1 195-206, 1938b

LORENTE DE NÓ, R *A Study of Nerve Physiology* New York Studies from Rockefeller Inst Med Res vols 132 and 133, p 384-477 1947

MacLEAN, P D The brain in relation to empathy and medical education *J Nerv Ment Dis* 144 374-382, 1967

MacLEAN, P D The limbic brain in relation to the psychoses In *Physiological Correlates of Emotion* edited by P Black New York Academic Press, chap 7 p 130-146 1970

MAGENDIE, F Experiences sur les fonctions des racines des nerfs rachidiens *J Physiol Exp Path* 2 276-279, 1822a

MAGENDIE, F Expériences sur les fonctions des racines des nerfs qui naissent de la moelle epinière *J Physiol Exp Path* 2 366-371 1822b

MAGNUS R Some results of studies in the physiology of posture *Lancet 2* 531-536, 585 588 1926

MAGOUN, H W An ascending reticular activating system in the brain stem *Arch Neurol Psychiat* 67 145-154 1952

MAGOUN, H W F HARRISON, J R BROBECK AND S W RANSON Activation of heat loss mechanisms by local heating of the brain *J Neurophysiol* 1 101-114 1938

MAGOUN, H W S W RANSON AND A HETHERINGTON Descending connections from the hypothalamus *Arch Neurol Psychiat* 39 1127-1149 1938

MAIRE, F W, AND H D PATTON Neural structures involved in the genesis of preoptic pulmonary edema' gastric erosions and behavior changes *Amer J Physiol* 184 345-350 1956

MARTIN, C Thermal adjustment of man and animals to external conditions *Lancet* 2 561-567, 617-620, 673-678, 1930

MASSION, J The mammalian red nucleus *Physiol Rev* 47 383 436, 1967

MATTHEWS, P B C Muscle spindles and their motor control *Physiol Rev* 44 219-288, 1964

MAURO, A The role of the voltaic pile in the Galvani Volta controversy concerning animal vs metallic electricity *J Hist Med* 24 140-150, 1969

McGINTY, D J, AND M B STERMAN Sleep suppression after basal forebrain lesions in the cat *Science* 160 1253-1255 1968

MILEDI R Properties of regenerating neuromuscular synapses in the frog *J Physiol (London)* 154 190-205, 1960

MILLER, N E Learning the visceral and glandular responses *Science* 163 434-445 1969

MITCHELL, J F, AND V SRINIVASAN Release of ^{3}H gamma aminobutyric acid from the brain during synaptic inhibition *Nature (London)* 224 663-666, 1969

MONIZ, E *Tentatives Operatoires dans le Traitement de Certaines Psychoses* Paris Masson et Cie 248 pp, 1936

MONNIER, M, AND L HÖSLI Humoral regulation of sleep and wakefulness by hypnogenic and activating dialysable factors In *Sleep Mechanisms*, edited by K Akert, C Bally, and J P Schade *Progr Brain Res* 18 118-123, 1965

MONTEMURRO, D G, AND J A F STEVENSON The localization of hypothalamic structures in the rat influencing water consumption *Yale J Biol Med* 28 396-403, 1955-1956

MOORE, D M, S F CHEUK, J D MORTON, R D BERLIN, AND W B WOOD, JR Studies on the pathogenesis of fever XVII Acti

vation of leucocytes for pyrogen production *J Exp Med* 131 179–188, 1970

MORRELI F Electrical signs of sensory coding In *The Neurosciences A Study Program* edited by G C Quarton, T Melnechuk and F O Schmitt New York Rockefeller Univ Press p 452–469 1967

MORUZZI G AND H W MAGOUN Brain stem reticular formation and activation of the E E G *Electroenceph Clin Neurophysiol* 1 455–473 1949

MOUNTCASTLE V B The problem of sensing and the neural coding of sensory events In *The Neurosciences A Study Program* edited by G C Quarton T Melnechuk and F O Schmitt New York Rockefeller Univ Press p 393–498 1967

MULLER G E, AND A PILZECKER Experimentelle Beitrage zur Lehre vom Gedachtniss *Z Psychol Physiol Sinnesorg* (suppl 1) 300 1900

MYERS R E Fuction of the corpus callosum in interocular transfer *Brain* 79 358 363 1959

NAKAYAMA T J S EISENMAN AND J D HARDY Single unit activity of anterior hypothalamus during local heating *Science* 134 560–561 1961

NASTUK W I, AND R L PARSONS Factors in the inactivation of postjunctional membrane receptors of frog skeletal muscle *J Gen Physiol* 56 218 249 1970

NAUTA, W J H Hypothalamic regulation of sleep in rats An experimental study *J Neurophysiol* 9 285–316 1946

NAUTA, W J H, AND D G WHITLOCK An anatomical analysis of the non specific thalamic projection system In *Brain Mechanisms and Consciousness*, edited by J F Delafresnaye Oxford Blackwell p 81 98 1954

OLDS J Hypothalamic substrates of reward *Physiol Rev* 42 554–604 1962

OLDS J, AND MILNER P Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain *J Comp Physiol Psychol* 47 419–427 1954

PAPEZ J W A proposed mechanism of emotion *Arch Neurol Psychiat* 38 725 743, 1937

PATON W D M E X VIZI AND M ABOO ZAR The mechanism of acetylcholine release from parasympathetic nerves *J Physiol (London)* 215 819 848 1971

PAVLOV, I P *Lectures on Conditioned Reflexes* translated by W H Gantt New York International Publ Co, 414 pp 1928

PEACHEY L D Transverse tubules in excitation contraction coupling *Fed Proc* 24 1124 1134, 1965

PENFIELD, W, AND H H JASPER *Epilepsy and the Functional Anatomy of the Human Brain* Boston Little Brown 896 pp, 1954

PENFIELD, W, AND T RASMUSSEN *The Cerebral Cortex of Man A Clinical Study of Localization of Function* New York Macmillan, 248 pp 1950

PENFIELD, W, AND L ROBERTS *Speech and Brain Mechanisms* New York Reprinted by Atheneum 286 pp, 1966

PENFIELD, W AND K WELCH The supplementary motor area of the cerebral cortex *Amer Med Ass Arch Neurol Psychiat* 66 289 317, 1951

PETERS, A S L PALAY AND H DE F WEBSTER *The Fine Structure of the Nervous System* New York Harper & Row, 198 pp, 1970

PRIBRAM, K H (ed) *Brain and Behavior 1 Mood, States and Mind Selected readings* Middlesex, England Penguin Books, 496 pp, 1969

PURPURA D P A neurohumoral mechanism of reticulo cortical activation *Amer J Physiol* 186 250–254 1956

PURPURA, D P Comparative physiology of dendrites In *The Neurosciences A Study Program* edited by G C Quarton T Melnechuk, and F O Schmitt New York Rockefeller Univ Press, p 372–393, 1967

QUARTON, G C, T MELNECHUK, AND F O SCHMITT (eds) *The Neurosciences A Study Program* New York Rockefeller Univ Press 962 pp, 1967

RAAB, D H, AND H W ADES Cortical and midbrain mediation of conditioned discrimination of acoustic intensities *Amer J Psychol* 59 59–83 1946

RAMÓN Y CAJAL, S Estructura del cerebelo *Gac Med (Catalana)* 11 449–457 1888

RAMÓN Y CAJAL S *Histologie du Systeme Nerveux de l'homme et des Vertebres* Paris Norbert Maloine, 2 vols 986 and 993 pp 1909 1911

RANSON, S W Some functions of the hypothalamus *Harvey Lect* 92–121, 1936 to 1937

RANSON S W AND P R BILLINGSLEY The superior cervical ganglion and the cervical portion of the sympathetic trunk *J Comp Neurol* 29 313–358 1918

RATNER S C AND M R DENNY *Comparative Psychology Research in Animal Behavior* Homewood Ill Dorsey Press 773 pp 1964

RENSHAW, B Activity in the simplest spinal reflex pathways *J Neurophysiol* 3 373–387 1940

RANSON S W AND S L CLARK *The Anatomy of the Nervous System*, Ed 10 Philadelphia W B Saunders, 622 p 1959

RENSHAW B Influence of discharge of motoneurons upon excitation of neighboring motoneurons *J Neurophysiol* 4 167–183, 1941

RENSHAW B Central effects of centripetal impulses in axons of spinal ventral roots *J Neurophysiol* 9 191–204 1946

RENSHAW B A FORBES AND B R MORISON Activity of iso cortex and hippocampus Electrical studies with micro electrodes *J Neurophysiol* 3 74 105 1940

RHEINBERGER M AND H JASPER Electrical activity of the cerebral cortex in the unanesthetized cat *Amer J Physiol* 119 186–196 1937

RICHTER, C P AND M HINES Production of grasp reflex in adult macaques by experimental frontal lobe lesions *Res Publ Ass Res Nerv Ment Dis* 13 211–224, 1934

RING G C The importance of the thyroid in maintaining an adequate production of heat during exposure to cold *Amer J Physiol* 137 582 588 1942

RITCHIE, A K AND A M GOLDBERG Vesicular and synaptoplasmic synthesis of acetylcholine *Science* 173 489–490, 1970

SASSIN, J F Neurological findings following short term sleep deprivation *Arch Neurol (Chicago)* 22 54–56, 1970

SCHMITT, F O (ed) *The Neurosciences Second Study Program* New York Rockefeller Univ Press, 1068 pp 1970

SCHNEDORF, G, AND A C IVY An examination of the hypnotoxin theory of sleep *Amer J Physiol* 125 491–505, 1939

SCHREINER, L AND A KLING Behavioral changes following rhinencephalic injury in cat *J Neurophysiol* 16 643–659, 1953

SCHRIER A M H F HARLOW AND F STOLLNITZ (eds) *Behavior of Nonhuman Primates* New York Academic Press, vol I 285 pp, 1965

SCOTT, J C, H C BAZETT AND G C MACKIE Climatic effects on cardiac output and the circulation in man *Amer J Physiol* 129 102–122, 1940

SEGUNDO, J P The reticular formation A survey *Acta Neurol Lat Amer* 2 245–281 1956

SHARE, L Control of plasma ADH titer in hemorrhage role of atrial and arterial receptors *Amer J Physiol* 215 1384–1389, 1968

SHEER, D E (ed) *Electrical Stimulation of the Brain* Austin Texas Univ Texas Press 641 pp, 1961

SHERRINGTON, C S *Integrative Action of the Nervous System* London Archibald Constable, 411 pp, 1906

SKINNER, B F *The behavior of Organisms An Experimental Analysis* New York Appleton Century Crofts 457 pp, 1938

SOLOMON R L, AND L H TURNER Discriminative classical conditioning in dogs paralyzed by curare can later control discriminative avoidance responses in the normal state *Psychol Rev* 69 202–219 1962

SPERRY, R W Cerebral organization and behavior *Science* 133 1749–1757, 1961

SPERRY, R W Some general aspects of interhemispheric integration In *Interhemispheric Relations and Cerebral Dominance*, edited by V B Mountcastle Baltimore Johns Hopkins Press, chap III, p 43–49, 1962

SPERRY, R W Split brain approach to learning problems In *The Neurosciences A Study Program* edited by G C Quarton, T Melnechuk, and F O Schmitt New York Rockefeller Univ Press, p 714–722, 1967

STEIN, D G, J J ROSEN, J GRAZIADEI, D MISHKIN, AND J J BRINK Central nervous system recovery of function *Science* 166 528–530, 1969

STERN K Note on the nucleus ruber magnocellularis and its efferent pathway in man *Brain* 61 284-289 1936

STOLWIJK J A J AND J D HARDY Temperature regulation in man—a theoretical study *Pflueger Arch* 291 129-162 1966

STROMBLAD B C, AND M NICKERSON Accumulation of epi nephrine and norepinephrine by some rat tissues *J Pharmacol Exp Ther* 134 154-159 1961

STRUMWASSER F Types of information stored in single neurons In *Invertebrate Nervous Systems* edited by C A G Wiersma Chi cago Univ Chicago Press chap 23 p 291-319 1967

TAUC L Transmission in invertebrate and vertebrate ganglia *Physiol Rev* 47 521-593, 1967

TEPPERMAN, J Horsley and Clarke a biographical medallion *Per spect Biol Med* 13 295-308 1970

TEPPERMAN J J R BROBECK AND C N H LONG A study of experimental hypothalamic obesity in the rat *Amer J Physiol* 133 P468-P469, 1941

THOMPSON R F, AND J A SHAW Behavioral correlates of evoked activity recorded from association areas of the cerebral cortex *J Comp Physiol Psychol* 60 329-339 1965

TOWER, S Pyramidal lesion in the monkey *Brain* 63 36-90 1940

TRACHTENBERG M C, AND D A POLLEN Neuroglia Biophys ical properties and physiologic function *Science* 167 1248-1252, 1970

TRUEX, R C, AND M B CARPENTER *Human Neuroanatomy* Baltimore Williams & Wilkins 673 pp 1969

UOTILA, U U On the role of the pituitary stalk in the regulation of the anterior pituitary with special reference to the thyrotropic hor mone *Endocrinology* 25 605-612 1939

VANE, J R Inhibition of prostaglandin synthesis as a mechanism of action for aspirin like drugs *Nature New Biol* 231 232-235 1971

VERHAART, W J C, AND M A KENNARD Corticofugal degenera tion following thermocoagulation of areas 4, 6 and 4 s in *Macaca mulatta J Anat* 74 239-254, 1940

VERNEY E G Antidiuretic hormone and the factors which deter mine its release *Proc Roy Soc (London) [Biol] Ser B* 135 25-106, 1947

VOGT C AND O VOGT Allgemeine Ergebnisse unserer Hirnfor schung Viete Mitteilung Die physiologische Bedeutung der archi tektonischen Rindenreizungen *J Psychol Neurol* 25 279-462, 1919

VON ECONOMO, C *Die Encephalitis lethargica* Berlin and Wien Urban and Schwarzenberg, 251 pp 1929

VON EULER U S Identification of the sympathomimetic ergone in adrenergic nerves of cattle (sympathin N) with laevo noradrenaline *Acta Physiol Scand* 16 63-74, 1948

VON EULER, U S Adrenergic neurotransmitter functions *Science*173 202-206, 1971

VON EULER, C, F HERRERO, AND I WEXLER Control mecha nisms determining rate and depth of respiratory movements *Resp Physiol* 10 93-108, 1970

VON FRISCH, K *The Dance Language and Orientation of Bees* Cam bridge, Mass Harvard Univ Press 566 pp, 1967

WALKER, A E Cerebral pedunculotomy for the relief of involuntary movements *J Nerv Ment Dis* 116 766-775 1952

WALSHE F M R AND E G ROBERTSON Observations upon the form and nature of 'grasping' movements and "tonic innervation seen in certain cases of lesions of the frontal lobe *Brain* 56 40-70 1933

WECHSLER, I S *A Textbook of Clinical Neurology* Ed 3 Philadel phia W B Saunders 826 pp, 1936

WEINSTEIN S AND H L TEUBER The role of preinjury education and intelligence level in intellectual loss after brain injury *J Comp Physiol Psychol* 50 535-539 1957

WHEATLEY M D The hypothalamus and affective behavior in cats A study of the effects of experimental lesions with anatomic corre lations *Arch Neurol Psychiat (Chicago)* 52 296-316 1944

WIERSMA C A G (ed) *Invertebrate Nervous Systems Their Sig nificance for Mammalian Neurophysiology* Chicago Univ Chicago Press 370 pp 1967

WILKINSON R T Loss of sleep *Proc Roy Soc Med* 62 903-904 1969

WILLIAMS M Memory studies in electric convulsion therapy *J Neurol Neurosurg Psychiat* 13 30-35, 1950

WILSON D M An approach to the problem of control of rhythmic behavior In *Invertebrate Nervous Systems* edited by C A G Wiersma Chicago Univ Chicago Press chap 17, p 219-229 1967

WINEGRAD S Role of intracellular calcium movements in excita tion contraction coupling in skeletal muscle *Fed Proc* 24 1146-1152 1965

WOLF A V *Thirst Physiology of the Urge to Drink and Problems of Water Lack* Springfield Ill Thomas 536 pp, 1958

WOODS, J W Behavior of chronic decerebrate rats *J Neurophysiol* 27 635-644 1964

WOODS J W AND P BARD Antidiuretic hormone secretion in the cat with a chronically denervated hypothalamus Copenhagen Proc Int Congr Endocr 1st p 113, 1960

WOOLSEY C N Organization of somatic sensory and motor areas of the cerebral cortex In *Biological and Biochemical Bases of Behav ior* edited by H F Harlow and C N Woolsey Madison Univ Wisconsin Press p 63-81, 1958

WURTMAN, R J Neuroendocrine transducer cells in mammals In *The Neurosciences* edited by F O Schmitt Second study pro gram New York Rockefeller Univ Press, p 530-538, 1970

YAMAMOTO, W S, AND J R BROBECK *Physiological Controls and Regulations* Philadelphia Saunders 362 pp, 1965

YAMAMOTO, W S, F S GRODINS L STARK, L D PAR TRIDGE, F E YATES, R D BRENNAN J URQUHART, AND H T MILHORN, JR Application of control systems theory to physi ology Physiology Society Symposium *Fed Proc* 28 46-88 1969

YOSHII, N, AND H OGURA Studies on the unit discharge of brain stem reticular formation in the cat I Changes of reticular unit dis charge following conditioning procedure *Med J Osaka Univ* 11 1-33, 1960

YOUNG, W C, R W GOY, AND C H PHOENIX Hormones and sexual behavior *Science* 143 212-218, 1964

INDEX

To state the obvious, this index covers all nine sections of BEST & TAYLOR'S PHYSIOLOGICAL BASIS OF MEDICAL PRACTICE, 9th ed What is not so obvious is the reason for inclusion of the complete index with each of the sectional print-outs The latter strategy will not only lead the reader to information in the print-out, but will also point out related and additional information in the other print-outs (or in the book itself)

α Adrenergic receptors
general properties, 9-64
mechanisms of action, 7-73
α Cells, pancreatic islets, 7-77
α Helix of protein, 1-124
α Rhythm, EEG, 9-107
α, β, δ, Chains of hemoglobin, 4-42
A band, myocardial sarcomere, 3-30 ff
A blood group, subgroups, 4-50
A nerve fibers, properties of, 1-42
Abdominal muscles in respiration, 6-49
Abdominal reflexes, 9 164
Abducens nerve (VI), 9 32
Aberration in eye, 8-67
ABO system of blood groups, 4 48
inheritance of, 4-50
Absolute refractory period
of muscle, 1 80
of nerve, 1 45
Absolute threshold, auditory, 8-31
Absorption, intestinal, 2-83 ff
of carbohydrate, 2 86
of electrolytes, 2 85
of fat, 2 91
of protein and amino acids, 2 89
of vitamins, 2-93
transamination during, 2 90
Absorption of air, 6 75
Absorption of water, 2-84
Acceleransstuffe, 9-60
Acceleration
angular, 8-41
coriolis, 8 54
Accessory muscles of respiration, 6-30
Accessory nerve, spinal, 9 38
Acclimatization to altitude, 6-66
Accommodation in bladder wall, 5 54
Accomodation in eye, 8-72
in depth perception, 8-118
Accommodation in nerve, 1-37
and activity of ion pumps, 1-50
Accommodation vs setpoint, 9-19
Acetate in lipid syathesis, 7-126
Acetazolamide vs carbonic anhydrase, 6-17
Acetoacetate, ketone body, 7-139
Acetone, ketone body, 7-139
Acetyl CoA
in control of ketogenesis, 7-139
in lipid synthesis, 7-124
significance in lipogenesis, 7-141
Acetyl CoA carboxylase, in control of lipogenesis, 7-142
Acetylcholine
as autonomic transmitter, 9-60
as neuromuscular transmitter, 1-57
as *Vagusstuffe*, 9-60
in control of muscle, 9-66
increase in coronary blood flow, 3-138
summary of transmitter sites, 1-58
Acetylcholine action
in adrenal medulla, 9-61
on epinephrine secretion, 7-72
on food intake, 2-3
on muscle membrane potential, 1-89
Acetylcholinesterase, in autonomic system, 9-61
Acetylene, arteriovenous, measurement of cardiac output, 3-159
Achalasia, 9-93
Achalasia and pain, 2-13
Acid-base balance, 5-27 ff
disturbed in dehydration, 4-117
disturbed in uremia, 5 46
various disturbances, 4-110
Acid-base properties of amino acids, 1-116
Acid hemolysis of RBC, 4-58
Acidemia, 4-107
Acidification of urine, 5 29 ff
Acidity and gastric emptying, 2-109
Acidity of gastric juice, 2-47
Acidophils of anterior pituitary gland, 7 8
Acidosis
as cause of dyspnea, 6 62
characteristics, 4 107
clinical types, 4-108
of renal dysfunction, 5 45
Acidosis, metabolic
causes, characteristics 4-108
in insulin deficiency, 7-87
Acidosis, respiratory, 4 107
Acromegaly, 7-143
ACTH (adrenocorticotrophic hormone)
amino acid sequence, 7-60
and aldosterone secretion, 7-63
and glucocorticoid synthesis, 7-60
ectopic production of, 7-64
measurement of, 7 61
ACTH-RF, 7 20
ACTH secretion
control of, 7-61
feedback control loops, 7-15
ACTH vs MSH, 7-61
Actin
arrangement in sarcomere, 3-30
concentration in myocardium, 3-32
interaction with tropomyosin and troponin, 3-38
Actin filaments, 1-62
Actin in muscle, 1-65
Actinins of muscle, 1-70
Actinomycin D
inhibits mRNA synthesis, 7 149
mechanism of inhibition of RNA synthesis, 1-171
Action currents in nerve, 1-43
Action potential, 1-37
history of theories, 1-45
modern concepts of, 1-46 ff
of auditory nerve fibers, 8-23
of skeletal muscle, 1-89
recording techniques, 9-10
sequence of ionic events, 1-48
technique for recording, 1-40
Action potential, compound
in sensory nerve, 8-8
of mixed nerve, 1-41
Action vs. resting potentials, 1-34
Activating mechanism, reticular, 9-140
Activation of amino acids, 1-173
Activation of EEG, 9-137
Active points in muscle, 1-85
Active state in muscle
effect on properties of muscle, 1-87
properties of, 1-82
Active transport, 1-21
Active transport in glands
bile secretion, 2-72
of I in thyroid, 7-30
saliva secretion, 2-21
Active transport in gut
amino acids, 2 89
electrolytes, 2-85
sugars, 2-87
Active transport in kidney
by renal tubules, 5-8 ff
of amino acids, 5-12
of glucose, 5-9 ff
of phosphate, 5-13
of sulfate, 5-13
of uric acid, 5-13
Active transport of cations, 1-5
energy cost in nerve, 1-52
Active vs inactive state of nerve, 1-51
Activity, suppressed by hypothalamus, 9-142
Actomyosin, 3 33
Acute renal failure, 5 43
Adaptation
in olfaction, 8-140
in vestibular system, 8-54
of nystagmus, 8-55
of touch sensation, 8-150
of urinary bladder, 5-53
renal, to acidosis, 5-32
Adaptation vs setpoint, 9-19
Addison's anemia, 4-33
Addison's disease
and renal dysfunction, 5 47
clinical features, 7-63
Adenine, 1-142
Adenohypophysis, 7-8 ff
actions of growth hormone, 7 143
control of hormone synthesis, 7-22
control of Leydig cells, 7-95
control of testis tubules, 7-99
control of thyroid function, 7-35
cytophysiology of, 7-8
embryology of, 7-24
hypothalamic connections, 9-41
listing of hormones, 7-8
mechanisms of control of, 7-8
mechanisms of hormone release, 7-21
neurohumoral control of, 7-13 ff
ADP (adenosine diphosphate), cardiac muscle contraction, 3-40
Adenosine triphosphate (ATP)
and fat absorption, 2-91
as purine derivative, 1-142
with G-actin, 1-66
Adenyl cyclase
activated by catecholamines, 7-73
in second messenger theory, 7-5
relation to lipolysis, 7-128

Adenylic acid,
biosynthesis of, 1-149
Adequate stimulus, 8-1
Adipocyte, 7-123
Adipose tissue, metabolism in, 7-123
Adipsia, hypothalamic, 9-126
Adolescence
and growth hormone, 7 143
growth of muscles, 7-146
Adrenal cortex, 7-50 ff
and renal dysfunction, 5-47
and water intoxication, 4-117
regeneration of, 7-50
Adrenal hormones (see name or type of hormone)
Adrenal medulla, 7-70 ff
disorders of, 7-75
Adrenal medullary catecholamines, regulation of heart, 3-115
Adrenal steroids
affect secretion of H^+, 5-34
affect sodium excretion, 5-25
in edema of cardiac failure, 5-45
Adrenaline (see Epinephrine)
Adrenergic nerves to gut, 2 8
Adrenergic system, 9-53
Adrenergic transmission, 9-62
Adrenochrome, 9-63
Adrenocortical dysfunction, 7-63
Adrenocortical tumors, 7-64
Adrenocorticotrophic hormone (see ACTH)
Aerobic metabolism of nerve, 1-50
Aeroembolism, 6-68
Afferent fiber types, 9-74
Afferent visceral nerves, 9-54
Afibrinogenemia, 4-80
After discharge, 9-72
After-image, colored, 8-114
Agglutination of RBC
by polyelectrolytes, 4-58
relation to blood groups, 4-48
Aggregation of ribosomes, 7-152
Aging and hearing, 8-35
Aging of lens, 8-68
Agnosia, 9-150
Agraphia, 9-151
Aids to hearing, 8 39
Air
absorption of, 6-75
composition, 6-5
Air conduction tests, 8-37
Airplane ascents, hypoxia in, 6-66
Airway during swallowing, 2-97
Airway resistance, 6-35
Alanine cycle, 7-136
Albumin of serum, 4-4
bilirubin binding, 2-65
binding to testosterone, 7-97
Albuminuria (see Proteinuria)
Aldolase activity, 7-127
Aldosterone
and renal tubular function, 7-59
biosynthesis of, 7-51
controls sodium excretion, 5-25
excess secretion of, 7-70
hypersecretion causes edema, 5-47
metabolic effects of, 7-59
metabolism of, 7-53
rate of renal excretion, 7-56
rate of secretion, 7-56
Aldosterone synthesis, regulation of, 7-62
Alkalemia, 4-107
Alkali reserve of plasma, 6-9
changes in dyspnea, 6-62
Alkalosis, 4-107
metabolic, 4-110
respiratory, 4-109
Allantoin from purines, 1-152
Alleles (of genes), 1-136
Allergy and asthma, 6-73
Allograft rejection, 4-99
All-or-none principle
in innervated muscle, 1-44
in nerve, 1-43
in skeletal muscle, 1-81
Allosteric control of enzymes, 7-114
Alloxan diabetes, 7-145
Alveolar air, 6-4, 5
gases of, 6-21
Alveolar air equation, 6-22
Alveolar gas exchange
anatomical shunts, 6-22
physiological shunts, 6-23
Po_2 gradient, 6-13
ventilation-perfusion abnormalities, 6-23
Alveolar vs total ventilation, 6-20
Amacrine cells of retina, 8-75
American population blood groups, 4-49
Aminoaciduria, 5-46
Aminoacidurias, pathogenesis of, 1-140
Amino acids, essential, 1-129
activation by tRNA, 1-173
active transport in kidney, 5-12
and ribosomal aggregation, 7-152
as precursors of glucose, 7-136
codon assignments for, 1-176
evoke insulin release, 7-79
for gluconeogenesis, 7-134
incorporation into protein, 7-152
intermediary metabolism of, 1-131 ff
intestinal absorption, 2-89
physicochemical properties of, 1-116
sources of renal NH_3, 5-31
summary of metabolic reactions, 1-135
utilization of, 1-129
Amino acid transport
affected by insulin, 7-84
control of entry into cells, 7-151
hormonal control of, 7-148
in plasma, 4-4
transmembrane, 1-139
Amino acids, 1-117
Amino acyl-tRNA, 1-180
p-Aminohippuric acid (PAH)
secretion by kidney, 5-14
tests based on clearance, 5-39
Ammonia
and acidification of urine, 5-31, 32
from deamination processes, 1-134
renal secretion of, 4-106
Amnesia, 9-157
5'-AMP and phosphorylase, 7-118
Amphetamine and food intake, 2-2
Ampulla, vestibular, 8-42
Amylase, pancreatic, 2-52
rate of secretion, 2-58
Amylase, salivary, 2-24
Amyloidosis of kidney, 5-40
Anaerobic muscle, 7-113
Anal orifice, defecation, 2-120
Anal reflex, 9-165
Analogue computation, neural, 9-13
Anaphylactic shock, 6-73
Anarthria, 9-150
Androgen (see also Testosterone)
Androgen action on fetal hypothalamus, 7-93
Androgen biosynthesis, 7-96
Androgen secretion in female, 7-106
Androgens from adrenal cortex, 7-51
Androgens from testis, 7-95
Anelectrotonus
and central computation, 9-13
in stimulation of nerve, 1-39
Anemia, 4-31 ff
absence of dyspnea, 6-61
coronary blood flow in, 3-139
cyanosis in, 6-80
hypoxia in, 6-64
Anemia, genetic, 4-20
Anemia of infants, 4-32
Angiotensin
and aldosterone synthesis, 7-62
controls Na^+ excretion, 5-25
increase in coronary blood flow, 3-139
vasoconstrictor effects of, 3-172
Angiotensin II, induces drinking, 9-126
Angular acceleration, 8-41
Animal vs human sensory studies, 8-4
Ankle jerk reflex, 9-86
Anosmia, 8-141
Anoxemia, 6-63
therapy with O_2, 6-81
Anoxia (see Hypoxia)
Anterior chamber of eye, 8-71
Anterior lobe of cerebellum, 9-43
Anterior lobe of pituitary gland (see also adenohypophysis), 7-8 ff
Anterior roots, spinal, 9 23
Anterior thalamic nuclei, 9-39
Antibiotics vs protein synthesis, 1-181
Antibodies
cellular, 4-90
humoral, 4-90
Antibodies of blood groups, 4-50
Antibodies to insulin, 7-86
Antibody responses, 4-91
Antibody synthesis, 4-98
Anticoagulants, typical, 4-81
Anticodons, 1-177
Antidiuresis, 5-18 ff
Antidiuretic hormone (ADH) (see also Vasopressin)
and renal permeability to water, 5-20
and water excretion, 5-18
control of release of, 9-126
Antidromic excitation, 9-76
Antridromic vs orthodromic stimulation, 1-53
Antigens for histocompatibility, 4-100
Antigens of ABO system, 4-48
Antigens, Rh in blood, 4-52
Antiglobulin test (Coombs), 4-50
Antigravity muscles
reflexes of, 9-69
tone of, 9-89
Antithrombin, 4-78
Antrum of stomach, 2-108
Anuria, 5-43
Aorta
blood flow during cardiac cycle, 3-105
pressure during cardiac cycle, 3-104
Aortic arch, innervation of, 3-177
Aortic body, 6-55
afferents in vagus nerve, 6-48
sensitivity to CO_2, 6-53
Aortic coarctation, cardiac circulatory effects of, 3-138
Aortic depressor nerve, control of blood pressure, 3-176
Aphasia, 9-148
Aplastic anemia, 4-19
Aplysia, conditioning in, 9-156
Apnea after hyperventilation, 6-41
in oxygen therapy, 6-82
Apneusis, 6-42

Apoferritin, 4-39
Apolar properties of protein, 1-126
Appetite, 2-1 ff
Appropriate stimulus, 8-1
Apraxia, 9-150
Aptyalism, 2-30
Aqueous humor, 8-71
Arabinose operon, 1-190
Archipallium, 9-104
Area 4, cerebral cortex, 9-112
Area 4 vs area 6, 9 115
Area 6, cerebral cortex, 9-116
Area 8, cerebral cortex, 9-117
saccadic eye movements from, 8 97
Argentaffin cells, intestinal, 2 80
Argyll-Robertson pupil, 9-166
Arterial pressure pulse, 3-194
Arteriolar nephrosclerosis, 5 41
Arteriosclerosis, in renal disease, 5 41
Artificial respiration, 6-83 ff
ASA 1951 curve (for hearing), 8-31
Ascorbic acid in adrenal cortex, 7 51
Asphyxia, 6-64
Aspirin, antipyretic, 9-136
Assembly of proteins, 1 183
Associated movement, automatic, 9-100
Association cortex, cerebral, 9 111
Astereognosis, 9-150
Asthma
bronchial, 6-72
cardiac, dyspnea in, 6-60
volume-flow relations, 6-34
Astrocytes, 1-31
Astronauts, ventilation of, 6-69
Atelectasis, 6-75
Athetosis, 9-170
ATP (adenosine triphosphate), cardiac muscle contraction, 3 40
ATP in muscle
and energy generation, 1 95
formation of cross-bridges, 1-73
ATPase
in adrenal medulla, 7 72
in membrane transport, 1-5
in slow vs fast muscle, 1 112
Atrial fibrillation, effect on volume output of heart, 3 118
Atrial mechanoreceptors, reflex to abdominal muscles, 6 50
Atrioventricular node, 3 56
block of, cause of arrhythmias, 3 75 ff
impaired conduction, 3-75
transmembrane potential, 3 49
Atrioventricular valves
closure, mechanism of, 3-106
movement during cardiac cycle, 3 105
Atrium, heart
anatomy, 3-25
blood volume in contractures, 3-98
conductance velocity, 3-55
contraction, 3 104
extrasystoles, 3-79
transmembrane potential, 3 48
vascular receptors, 3-186
weight, 3-26
Atropine, effect on gastric secretion, 2-46
Audiogram, von Békésy, 8-37
Audiometric zero curve, 8-31
Audiometry, 8-35
Audiometry for speech, 8-38
Audition (see Hearing)
Auditory cortex, 8-30
Auditory discrimination, 8-30
Auditory hair cells, 8-21
potentials of, 8-19
Auditory nerve, 8-15
Auditory system
course of afferent fibers, 8-25
efferent fibers in, 8 25
information processing, 8-29
lateralization in, 8-26
Auditosensory area, 9-111
Auerbach's plexus, esophageal, 2-103
Auscultatory gap, blood pressure measurements, 3-151
Autoimmune thyroiditis, 7 41
Automatic associated movement, 9-100
Automaticity of respiratory centers, 6-44
Autonomic blockade of gastric secretion, 2 47
Autonomic centers, medullary, 9-65
Autonomic conditioning, 9 146
Autonomic control
of gastric secretion, 2 35
of pancreas, 2-55 ff
of salivary secretion, 2 27
Autonomic nervous system, 9-53 ff
higher controls, 9 64
to digestive tract, 2 8 ff
to salivary glands, 2 19
Autophagia in lysosomes, 1-11
Autoradiography of chromosome, 1 165
Avogadro's law of gases, 6-1
Avoidance conditioning, 9-153
Axon hillock, 1-29
Axon of nerve, 1-30
Axonal transport to neural lobe, 7-24

β Adrenergic receptors, 7-73
in autonomic system, 9-64
β Cell, pancreatic islets, 7 77
defect in diabetes mellitus, 7-87
β-Hydroxybutyrate, 7-139
β Lipoprotein, 2-92
β-Pleated sheet of protein, 1-124
3β-ol Deficiency, 7-69
B nerve fibers, properties of, 1-42
Babinski's sign, 9-164
Balance, glomerulotubular, 5-12
Barany chair, 8-49
in clinical tests, 8-55
Baroreceptors
and vasopressin release, 7-27
influence on respiration, 6-48
reflex to abdominal muscles, 6-50
Barospirator of Thunberg, 6-84
Barr body, sex chromatin, 7-91
Basal ganglia, cerebral, 9-99 ff
anatomy and relations, 9-49 ff
summary of diseases, 9-168
Base composition of DNA, 1-158
Base pairing in DNA, 1-161
Base pairs, models of, 1-152
Base sequence of DNA, 1-161
Basic electrical rhythm, gastric, 2-106
Basic taste sensations, 8-121
Basic visual functions, 8-98 ff
Basilar membrane of cochlea, 8-14
Basophils of anterior pituitary gland, 7-8
Basophils of blood, 4-62
Bat wing capillaries, 4-84
Bayliss mechanism in blood circulation, 3-165
Behavior
after visual deprivation, 8-95
in study of sensation, 8-4
Belching, 2-104
Bell-Magendie law, 9–23
Bell's palsy, 9–37
Bends, 6-68
Bernard, Claude
and salivary secretion, 2-26
milieu interne, 9-58
sucre piqûre, 9-65
Bernstein's theory of action potential, 1-45
Betz cells, 9-27
Bezold-Jarisch reflex, 3-185
Bicarbonate-carbon dioxide ratio
in acidosis, alkalosis, 4-107
normal values, 4-105
Bicarbonate ions
and carbonic acid buffer system, 4-104
and CO_2 transport in plasma, 6-16 ff
and gastric secretion, 2-39
in pancreatic juice, 2-50
rate of pancreatic secretion, 2-58
renal transport, 5-27 ff
Biceps jerk, 9-87
Bile, 2-60 ff
electrolytes in, 2-60
essential for cholesterol absorption, 2-93
in gallbladder, 2-73
miscellaneous constituents, 2-69
osmolarity, 2-61
Bile acids, conjugation of, 2-62
Bile ducts
anatomy of, 2-72
obstruction of, 2-68
pressures inside, 2-75
Bile pigments, 2-64 ff
conjugation of, 2-66
enterohepatic circulation of, 2-68
Bile salts, 2-61 ff
enterohepatic circulation, 2-63
excretion of, 2-64
Bile secretion
cellular mechanisms, 2-71
control of, 2-70
rate of, 2-61
Bilirubin, 2 64
conjugation of, 2-66
from hemoglobin degradation, 4-35
hepatic excretion, 2-67
sites of production, 4 42
Biliverdin, 2-64
Binaural hearing, 8-28
Binaural sound localization, 8-33
Binding assay for nucleotide sequence, 1–176
Binocular retinal disparity, 8-120
Binocular vision, coding in cortex, 8-94
Bioelectricity, principles of, 1-32 ff
Biosynthesis of adrenal hormones, 7-50
Biosynthesis of RNA, 1-169
Biot's breathing, 6-58
Bipolar cells of retina, 8-75
Birth
changes in circulation, 6-76
resuscitation of infant, 6-85
Bitter taste, 8-122
Bladder, urinary, 5-47 ff
electron micrograph of, 1-4
physiological capacity, 5-53
reflexes in spinal shock, 9-163
Bleeding time, 4-78
Blind spot, 8-70
Blocking of thyroid hormone synthesis, 7-32
Blood, 4-1 ff
capillary exchange of, 3-16
rheological properties of, 4-61
vascular beds, patterns of, 3-5
vascular distribution, 3-5

Blood cells
granulocytes, 4 62 ff
in blood stored for transfusion, 4-55
model of origin, 4 11
sites of formation, 4-11
Blood clot retraction, 4-75
Blood clotting, 4-75 ff
abnormalities of, 4-79 ff
enhancement of, 4-82
prevention of, 4-81
Blood components for transfusions, 4-53
Blood constituent (see name of constituent)
Blood creatinine, 5-38
Blood flow
aortic phasic, 3-196
arterial, 3-196 ff
bronchial, 3 205
capillary, 3-198
cerebrum, 3-211 ff
circulation time, 3-162
control of, and regulation, 9-124
cutaneous, and temperature regulation, 9-129
electromagnetic flow meter in measurement of, 3 156
exercise effects on, 3-208
fetal, compared with adult, 3-239
Fick principle in measurement of, 3-158
foot, 3 214
hand, 3-214
indicator dilution principle, 3-161
kidney, 3 160, 3-228 ff
liver, 3-226
lungs, 3-204
measurement of, 3-152 ff
measurement of, to kidney, 5-15
mechanical devices for assistance of, 3-153
mesenteries, 3 224
pendulum flow meters in measurement of, 3-156
perfusion systems for measurement of, 3-153 ff
peripheral arteries, 3 197
phasic, flow meters in measurement of, 3-156
regulation, 3-5
respiratory effects on, 3 207
rotameter measurement of, 3-155
skeletal muscle, 3-217
skin, 3-213 ff
sonar flow meter in measurement of, 3-156
spleen, 3-225
to salivary glands, 2-20
vascular resistance in lungs, 3-205
vasomotor control in lungs, 3-209
velocity, 3-4
venous drainage in measurement of, 3-153
vessel velocity changes, 3-18
visualization of microcirculation, 3-152
Blood gases (see also Carbon dioxide, Oxygen, Carbon monoxide)
measurement of, 6-8 ff
transport of, 6-5 ff
Blood glucose, action of insulin, 7-83
Blood glucose and feeding, 2-4
Blood group antigens, 4-49
Blood groups, 4-48 ff
antibodies of, 4-50
miscellaneous types, 4-53
Blood nonprotein nitrogen, 5-38
Blood platelets, 4-73 ff
in thrombosis, 4-81
Blood pressure
auscultatory measurement of, 3 150 151
autoregulation of, 3-6
average value, according to age, 3-189
cardiac output and, 3-20, 3-192
diastolic, 3-190
digestion and, 3-190
distribution in cardiovascular system, 3 6
effects of pH variation, 3-183
emotions and, 3-190
exercise effects on, 3-190, 3-202 ff
extrinsic control, 3-8
elasticity of vessel walls as determinant of, 3-193
flush method for measurement of, 3-152
gauges for measurement of, 3-149
gradient, species comparison, 3-4
heart pumping action as determinant of, 3-191
human aorta, effect of aging, 3-14
indirect measurement in animals, 3-152
intrinsic control, 3-5
level of critical closure of vessels, 3 194 ff
measurement of, 3-148 ff
microphones in measurement of, 3 152
nonoscillatory, 3-148
peripheral resistance unit, 3 192
postural effects on, 3-190, 3-202 ff
pulmonary vessels, 3-21
pulse detectors in measurement of, 3 152
regulation, 3-5
relation of colloid osmotic pressure and, 3-4
respiratory effects on, 3-302, 3-207
systolic, 3-190
transfer function, 3-8
transfusion effects, 3 18
ultrasound in measurement of, 3 152
vascular peripheral resistance as determinant of, 3-191 ff
vein capacity as determinant of, 3 200
venal caval flow as determinant of, 3-201
ventricular contraction as determinant of, 3-200
viscosity of blood as determinant of, 3-193
Blood pressure, arterial, 3-4, 3-189, 3-194, ff
and chemoreceptive reflexes, 6-55
influence on respiration, 6-49
Blood pressure, arteriolar, 3-15
Blood pressure, capillary, 3-15, 3-198 and glomerular filtration, 5-5
Blood pressure, cardiovascular, 3-2
Blood pressure, venous, 3-17, 3 198 ff
Blood sedimentation, 4-60
Blood supply to kidney, 5-2
Blood transfusion, 4-48 ff
incompatible transfusion, 4-53
Blood urea, 5-36 ff
Blood velocity, measurement of, 3-161 ff
Blood vessels
α-adrenergic receptor vasoconstriction of, 3-171
β-adrenergic receptor vasodilatation of, 3 171
diameters and blood content, 3-13
local vs remote control of vasoconstriction, 3-173 ff
of retina, 8 72
parasympathetic vasodilator fibers, 3-170
reciprocal vasoconstriction reactions, 3 180
reflex vasoconstriction, 3-176
remote vs local control of vasoconstriction, 3-173 ff
smooth muscle constriction of, 3-164
vascular tone, 3-172, 173
vasconstrictor fibers of, 3-166, 3-167
vasodilator fibers of, 3-170
Blood viscosity, 4 60
Blood volume
abnormalities of, 4-119
arterioles, 3-15
capillaries, 3-15
clinical features of changes, 4 120
distribution of, 3 2
dog circulatory system, 3 6
normal values, 4-119
pulmonary vessels, 3-21
regulation, 3 5
transfusion effects, 3-18
varies with ambient temperature, 9-130
veins, 3 17
Blood volume estimations, 4-118 ff
Blood volume regulation, 4 117 ff
Body fluid exchange, 4 113 ff
Body temperature regulation (see Regulation of body temperature)
Body weight, control of, 9 127
Bohr effect on hemoglobin, 4 43
Bohr integration, O_2 diffusing capacity, 6 25
Bolus, swallowing, of, 2-97
Bonds in protein molecules, 1-122
Bone and calcium exchanges, 7 44
Bone conduction tests, 8-37
Bone growth, hormonal control, 7-144
Bone marrow
and hemopoiesis, 4-11
drug effects on, 4-19
forms bile pigments, 2-65
granulocytopoiesis, 4-64
in erythropoiesis, 4-21
lymphocytopoiesis, 4 95
suppressed by drugs, etc , 4 71
Border contrasts, visual, 8-90
Boyle's law of gases, 6-1
Brachium conjunctivum, 9-47
Brachium pontis, 9-47
Bradykinin
and inflammation, 4-64
in saliva, 2-25
vasodilation effect in skeletal muscle, 3-219
Brain
embryology of, 9-20
stimulation of, 9 8
Brainstem
auditory systems of, 8-26
autonomic outflow, 9-57
centers for regulation, 9-125
connections of taste fibers, 8 128
control of respiration, schema, 6-46
neural tracts, 9-24 ff
respiratory centers, 6-42 ff
single vestibular units, 8-58
structure, 9 30
vestibular nuclei, 8-49
Breast development, in pregnancy and lactation, 7-109
Breathing (see also Pulmonary ventilation, Respiration)
abnormal patterns, 6-58
at high pressures, 6-83
changes in pneumonia, 6-71
control of, 6-41 ff
efficiency of, 6-37
mechanics of, 6-28 ff
periodic, 6-58

positive pressure, 6-49
work of, 6-36
Brightness discrimination, 8-99
Brodmann's areas of cerebral cortex, 9-105
Bromide ion in saliva, 2-23
Bromosulphalein
excretion by liver, 2 72
use in measurement of kidney blood flow, 3-160
Bronchioles in asthma, 6-72
Bronchopneumonia, 6-69
Brunner's glands, 2-80
Brush border, intestinal, 2 87
Buffer action and pH of urine, 5-30
Buffer systems for H^+, 4-104 ff
Bulbar facilitatory mechanism, 9 93
Bulbar inhibitory mechanism, 9-93

C cells of thyroid gland, 7 46
C nerve fibers, 1-42
Caffein on gastric secretion, 2-46
Caisson disease, 6-68
Cajal, Ramon y
neuron doctrine, 9 68
staining of neurons, 9-44
Calcitonin, 7-42 ff
action on plasma calcium, 7-48
chemical structure, 7 47
comparative physiology of, 7-47
from C cells of thyroid, 7-46
Calcium
hypercalciuria, 5-42
in control of muscle, 9 66
intestinal absorption, 2 85
Calcium homeostasis, 7 42 ff
Calcium ions
and hormone secretion, 7 21
and insulin release, 7 80
association with G-actin, 1-66
bond formation in muscle, 1-78
cardiac excitation contraction, 3 36 ff
in blood clotting, 4-75
in epinephrine release, 7-72
in excitation-contraction coupling, 1 90
on neuromuscular transmission, 1-57
role in action potential, 1-50
Calcium of plasma
effect of calcitonin, 7-48
effect of PTH, 7-44
removal for anticoagulation, 4-81
transport of, 7-43
Caloric stimulation of vestibule, 8-49
Calorigenic action of epinephrine, 9-136
Canals, semicircular, 8-41
Cancer and immune response, 4-100
Cannon, Walter Bradford, homeostasis, 9-58
Capacitance of cell membrane, 1-15
Capacitance of muscle membrane, 1-91
Capacity of bladder, 5-53
Capillaries
abnormalities in kidney, 5-41
and lymph formation, 4-83
blood O_2 in cyanosis, 6-79
exchange of O_2, CO_2, 6-17
of renal glomeruli, 5-4
Starling principle, 4-122
Capillaries, lymphatic, 4-83
Capillary membrane
blood exchange in, 3-16
permeability of, 4-122
Carbaminohemoglobin, 6-15
Carbohydrate metabolism
and corticosteroid actions, 7-57
control of in muscle, 7-112 ff
Carbohydrates, intestinal absorption, 2-86
Carbon dioxide
arteriovenous, in measurement of cardiac output, 3-158
blood dissociation curves, 6-16
cerebral circulation response to, 3-212
chemodetection, central, 6-47
chemoreceptor effect on circulation, 3-182 ff
content in blood, 6-6
effect on pulmonary ventilation, 6-52
exchange in alveoli, 6 4
in acidosis, alkalosis, 4-108
interactions with hemoglobin, 4-46
measurement in blood, 6-8 ff
resistance to diffusion in lung, 6 27
retention in asphyxia, 6-64
retention in pneumonia, 6-69
sensitivity of chemodetectors 6-53
used in O_2 therapy, 6-82
used in resuscitation, 6-85
Carbon dioxide electrode, 6 10
Carbon dioxide narcosis, 4-108
Carbon dioxide tension
and acidification of urine, 5-34
and epileptic seizures, 9 172
and multiple factor theory, 6-56
and signal generation, 9-16
in alveoli of lung, 6-22
in blood, 6-5
in emphysema, 6-74
in tissues, 6 19
influence on breathing, 6-41
on hemoglobin oxygen binding, 4 43
Carbon dioxide transport, 6-1 ff
and buffering of H^+, 4-106
as carbaminohemoglobin, 6-15
at high altitude, 6-67
chloride-bicarbonate shift, 6-18
effect of oxygen, 6-15 ff
Carbon monoxide
binding with hemoglobin, 4-46
diffusing capacity in lung, 6-26
effect of O_2 in blood, 6-15
O_2 used in poisoning, 6 83
poisoning from, 6-78
Carbonic acid
and bicarbonate buffer system, 4 104
as source of hydrogen ions, 4-103
Carbonic acid-bicarbonate system, in renal transport of H^+, 5 27
Carbonic anhydrase, 6-16 ff
and acidification of urine, 5-29 ff
in gastric juice, 2 38
in parietal cells, 2-39
inhibited by cyanide, 6-78
inhibitors depress Na^+ transport 5 26
Cardia of stomach, 2-105
Cardiac asthma, 6 60
Cardiac cycle, 3-101 ff
Cardiac depressor nerve, control of blood pressure 3-176
Cardiac dyspnea, 6-59
Cardiac edema
and renal dysfunction, 5-44
increased extracellular fluid, 4-88
Cardiac output, at high altitude, 6-67
Cardiac reflexes, 3 114, 115
Cardiospasm, 2-109
Carotid body
afferents in vagus nerve, 6-48
properties of, 6-55
sensitivity to CO_2, 6-53
Carotid sinus
baroreceptor response, 3-179
blood pressure control mechanism, 3-177
innervation of, 3-177
reflex arc, 3-178
Carrier in active transport, 1-24
Carrier-mediated transport
mechanisms and models, 1 20
of glucose, 7-114
of glucose, renal, 5-9 ff
Carrier molecules, defects of, 1-140
Casein, digestion by rennin, 2-37
Catabolite repression, 1 190
Cataplexy, 9-168
Catecholamines
and autonomic transmission, 9-62
assay for, 7-72
biosynthesis, autonomic, 9 62
conjugation and inactivation, 9-63
control gonadotrophin secretion, 7 15
details of biosynthesis, 7-71
details of metabolism, 7-72
effect on food intake, 2-3
heart rate regulation during exercise, 3-122
physiological actions, 7-73
storage and release, 7-72
summary of metabolic actions, 7 75
Catechol-O-methyl transferase, 7 72
Catelectrotonus
and central computation, 9-13
in stimulation of nerve, 1-39
Cathode ray oscilloscope, 1-40
Cathode, stimulation of nerve, 1 38
Caudate nucleus, 9-50
Cell columns of spinal cord, 9-21 ff
Cell cycle of hemopoiesis, 4 13
Cell division as growth process, 7-146
Cell indices, erythrocytes, 4-31
Cell mediated immunity, 4-101
Cell membrane (see Membrane, cell)
Cell organelles, 1-2
Cell physiology (see also Membrane, names of cell structures and processes), 1-1 ff
Cell renewal system, 4-10
granulocytopoiesis, 4-65
in erythropoiesis, 4-21
Cell surface and rate of transport, 1-6
Cell types of adenohypophysis, 7-8
Cellular immunity and graft rejection, 4-98
Center
reflex, for defecation, 2-120
swallowing, 2-96
vomiting, 2-110
Centers, autonomic, medullary, 9-65
Central excitatory state (CES), 9-13 ff
of motor neuron pool, 9-71
Central inhibitory state (CIS), 9-13 ff
Central nervous system (see Brain)
Central synaptic delay, 9-75
Centrioles, 1-13
Centrosome, absent in neuron, 1-30
Cerebellar-cerebral connections, 9 47
Cerebellar cortex, 9-44
Cerebellar disease, 9-168
Cerebellar nuclei, 9-46
Cerebellum
comparative physiology of, 9-97
connections of, 9-46 ff
control of movement, 9-96
control of posture, 9-97
experimental study, 9-49
flocculonodular lobe, 9-98
neuronal interrelations, 9-48
neuronal organization, 9-44
signs of disease, 9-168
structure and divisions, 9-42 ff
vestibular input, 8-49
Cerebral-cerebellar connections, 9-47

Cerebral cortex, 9-104 ff
association areas, 9-111
comparative physiology, 9-88
comparative physiology of deficits, 9-113
control of vasoconstriction, 3 170
decortication, 9-94
dominant hemisphere, 9-120
electrical activity, 9-105
fine structure, 9 104
lateralization of function, 9 119
localization of function, 9 104
pre- and orbitofrontal regions, 9-117
Cerebral cortex, motor, 9-112 ff
area 4 vs area 6, 9-115
autonomic connections, 9 65
frontal eye fields, 9-117
in control of respiration, 6-46
projections from, 9-89
speech mechanisms, 9-148
visceral motor areas, 9-116
Cerebral cortex, sensory, 9-109
auditory connections, 8-26
olfactory connections of, 8-138
single cell activity, visual, 8-93
thalamocortical fibers, 9 108
visual areas, 8-78
Cerebrospinal fluid (CSF)
affects chemosensitivity, 6-47
hypnotoxin in, 9-142
Ceruloplasmin, 4-6
Cerveau isole, 9-138
Cervical sympathetic chain, 9-55
Chalasia, 2-103
Charles' law of gases, 6-2
Chemical effects
on hemolysis, 4-58
on iodide trapping, 7 30
on secretion of saliva, 2-29
Chemical senses (see Taste, Olfaction)
Chemical sensitivity in gut, 2-11
Chemical transmission
adrenergic, autonomic, 9-62
cholinergic, autonomic, 9-60
history of, 1-52
in autonomic system, 9-59
in basal ganglia, 9-100
in brain and spinal cord, 9-14
mechanism of, 1-57
Chemical transmitters
distribution in CNS, 9 83
storage and release, 1-58
Chemical vs electrical transmission, 9 82
Chemoreceptors
central, for CO_2, H^+, 6-47
CO_2 sensitivity, 6-53
influence on abdominal muscles, 6-50
sensitivity to H^+, 6-56
sensitivity to low O_2, 6-54
Chemotaxis of leukocytes, 4-63
Chemotherapy of leukemia, 4-19
Chest in emphysema, 6-73
Chest in ventilation, 6-32 ff
Chewing, 2-95
Cheyne-Stokes breathing, 6-58
Chief cells of stomach, 2-31
Childbirth and oxytocin, 7-28
Chloride-bicarbonate shift in blood, 6-18
Chloride current in muscle, 1-92
Chloride ions
and Hodgkin-Huxley theory, 1-36
and spinal EPSP, 9-78
and synaptic potentials, 1-55
in countercurrent system, 5-20
renal transport of, 5-23 ff
Chloride pump, gastric, 2-36
Cholagogues, 2-70
Cholecystectomy, 2-75
Cholecystokinin, 2-15
Cholelithiasis, 2-77 ff
Choleretics, 2-70
Cholesterol
and cholelithiasis, 2-77
constituent of bile, 2-69
in bile acid synthesis, 2-62
in corticosteroid biosynthesis, 7-52
in glucocorticoid synthesis, 7 60
intestinal absorption of, 2-93
plasma levels of, 4-8
Cholic acid in bile, 2 62
Cholic acid in cholelithiasis, 2-78
Choline acetylase, 9-60
Cholinergic nerves to gut, 2-8
Cholinergic system, 9-53
Cholinergic transmission, 9-60 ff
Cholinesterase, in autonomic system, 9-61
Chorda tympani nerve
taste fibers in, 8-126
to salivary glands, 2-19
Chorda tympani syndrome, 2-30
Chorea
as rhythmic mechanism, 9-102
types and symptoms, 9 169
Choreoathetosis, in thalamic lesions, 9-167
Chorionic gonadotropin (HCG), 7 108
Chromaffin cells of adrenal medulla, 7-71
Chromatic aberration, 8-67
Chromatin and nuclear sex, 7-88
Chromatin, nuclear, 1 7
Chromatography
of blood gases, 6-10
of gas mixtures, 6-8
of releasing factors, 7-18
Chromatolysis of Nissl bodies, 1-29
Chromophobes of anterior pituitary gland, 7-8
Chromosomal sex, 7-88
Chromosome mapping, 1-162
Chromosomes
chemical nature of, 1-7
in eukaryotic cells, 1-193
of stem cells in hemopoiesis, 4-18
Chromosomes X and Y,
and sex differentiation, 7 88
morphology of, 1-137
Chronaxie, 1-38
Chylomicrons
and fat absorption, 2-91
of lymph, 4-85
Chymotrypsin, 2 52
Chymotrypsinogen, 2-52
Cilia and flagella, 1-13
Cilia in cochlea, 8-19
Cilia in olfactory receptor, 8-135
Circulation, response to thermal stress, 9-130
Circulation, control of, as example of regulation, 9-124
Citrate, in lipid synthesis, 7-124
Clasp-knife reflex, 9-73
and muscle spindles, 9-85
Clearance
of creatinine, 5-39
of free water, 5-22
of inulin, tests based on, 5-39
of PAH, tests based on, 5-39
of urea, 5-35, 39
Clearance methods
for glomerular filtration, 5-6 ff
for study of kidney, 5-3
Clearing factor lipase, 7-128
Climbing fibers of cerebellum, 9-44
Clonal selection theory of antibodies, 4 91
Clones from marrow, 4 18
Clonus
as rhythmic mechanism, 9-103
as stretch reflex, 9-87
Clot dissolution, 4-77
Clot retraction time, 4-78
Clotting mechanism, 4-75 ff
Clothing and temperature regulation, 9-130
CNS control of respiration, 6-52 ff
CNS control of urinary tract, 5-51
Coagulation factors in lymph, 4-85
Coagulation of blood (see also Blood clothing), 4-75 ff
Coagulation time, 4-78
Cobalt in vitamin B_{12}, 4-38
Cochlea, 8-14 ff
electrical potentials in, 8-22
innervation of, 8-23
microscopic anatomy, 8-16
Cochlea vs vestibular system, 8-15
Cochlear efferent system, 8-25
Cochlear microphonic potentials, 8-20
Cochlear nuclei in hearing, 8-26
Code, genetic, 1-175 ff
Coding, neural, 9-12
for color vision, 8-91
of sensory response, 8-9
of sensory signals, 8-3
Codon-anticodon interaction, 1-177
Codons for protein synthesis, 1-175
Coefficient of O_2 utilization, 6-14
Coenzyme A from pyruvate oxidation, 1-103
Cold agglutinin and hemoglobinuria, 4-59
Cold receptors, 8-153
Colitis, ulcerative, 2-117
Colligative properties of water, 5-17
Colloid of thyroid gland, 7-33
Colon, motility of, 2-117
Colonic motility, abnormalities, 2-119
Colonic secretion, 2-82
Color anomaly, 8-117
Color blindness, 8-114 ff
Color discrimination, 8-113
Color vision, 8-113 ff
and stimulus wavelength, 8-91
coding in cortical units, 8-94
coding in lateral geniculate body, 8-91
sensitivity of receptors, 8-83
Columns of spinal cord, 9-21 ff
Commissures of brain, 9-119
Commissures of spinal cord, 9-22
Common bile duct, 2-72
Communication and speech, 9-147
Comparative physiology
of blood, 4-1
of calcitonin, 7-47
of cerebellum, 9-97
of cerebral cortex, 9-88
of deficits from cerebral cortex, 9-113
of insulin, 7-76
of membranes, 1-26
of releasing factors, 7-20
of sensory processes, 8-4
of temperature regulation, 9-128
of velocity of nerve impulse, 1-43
Compartment analysis
for erythropoiesis, 4-22
in hemopoiesis, 4-16
of granulocytopoiesis, 4-65
Compensation for acid-base disturbances, 4-107 ff
Compensation, intracodon, 1-185

Competition in carrier transport, 1-20
Competition of sugars for carrier, 7-115
Compliance of lung, 6-31
Compound action potential
in sensory nerve 8-8
of mixed nerve, 1-41
Compromise in control systems, 9-17
Computer analysis of sensation, 8-6
Computer for clinical diagnosis, 4-112
Computer print out, H^+ disturbance, 4-111
Concentration gradient and transmembrane movement, 1-18
Concentration potentials, 1-33
Concentration test, urine, 5-39
Conditioned reflexes, 9-151 ff
Conditioned secretion of saliva, 2-26
Conditioning, autonomic, 9-156
Conditioning, classical, 9-151
electrical recordings 9-151
Conditioning, instrumental, 9-153
Conditioning, mechanism of, 9-156
Conditioning, neuronal, 9-156
Conduction deafness, 8-34
Conduction velocity vs fiber size, 1-42
Cone pigments and color vision, 8-91
Cones vs rods
distribution of, 8-69
in dark adaptation, 8-101
Conformation of protein, 1-127
Congenital adrenal hyperplasia, 7-66
Congenital cardiac defects, 6-77
Congenital neutropenia, 4-72
Conjugation of bacteria, 1-163
Conjugation of bile pigments, 2-66
Connectors, neural, 9-2
Consciousness of rotation, 8-53
Consciousness of sensation, 8-2
Constipation, 2-120
Continence of urine, 5-54
Contraction of muscle, 1-61 ff
isometric, 1-82
isotonic, 1-79
Control
of energy balance, 9-127
of enzymes in muscle, 7-114
of exocrine glands, 9-53 ff
of intestinal motility, 2-114
of skeletal muscle, 9-66 ff
of smooth muscle, 9-53 ff
of water balance, 9-125
Control, endocrine
of cell size and number, 7-146
of gluconeogenesis, 7-137
of lipolysis, 7-130
of protein synthesis, 7-153
vs neural control, 9-121
Control, humoral
of sleep and wakefulness, 9-142
of thermal exchange, 9-136
Control, neural
hierarchy of levels, 9-15
local processes, 9-11
of locomotion, 9-96
of movement, 9-96
of movement, cortical control, 9-112
of posture, 9-88
of skeletal muscle, 9-66 ff
of swallowing, 2-96
of sweating, 9-131
of thermal exchange, 9-128 ff
of urinary tract, 5-50 ff
of ventilation (schema), 6-46
priority and compromise, 9-17
resume of deficits, 9-161 ff
setpoint of, 9-19
vs endocrine control, 9-121
Control points in purine synthesis, 1-150
Control systems in genetics, 1-136
Control systems, summary of, 9-121 ff
Control, transmission of, 9-11
Control vs regulation, 9-18
Convergence
in depth perception, 8-119
in spinal cord, 9-71
Convulsions
from irritation of area 4, 9-115
in water intoxication, 4-117
origin of, in epilepsy, 9-172
Coombs test, 4-50
Copper injures basal ganglia, 9-99
Copper transport in plasma, 4-6
Coproporphyrins, 4-42
Core enzyme of polymerase, 1-170
Cori cycle of lactic acid, 7-135
Coriolis acceleration, 8-54
Cornea, 8-70
Coronary artery
blood flow, 3-131 ff
collateral circulation, 3-143
drugs vs collateral circulation in occlusion of, 3-144, 145
stimulation of collateral circulation, 3-145
Corpus callosum, function of, 9-119
Corpus luteum, 7-101
Cortex (see Adrenal cortex, Cerebellar cortex Cerebral cortex, etc)
Corti, organ of, 8-17
Cortical inhibitory mechanism, 9-145
Corticipetal fibers, 9-52
Corticobulbar tract, 9-27
Corticofugal fibers, 9-51
Corticospinal tract, 9-27
origin, areas 4 and 6, 9-113
Corticosteroid biosynthesis, 7-50
defect in congenital adrenal hyperplasia, 7-67
scheme of pathways for, 7-52
Corticosteroids (see also glucocorticosteroids)
assay of, 7-53
metabolism of, 7-53
physiological effects, 7-56
Corticotrophin releasing factor (CRF), 7-20
Cortisol (see also glucocorticosteroids)
actions of, 7-57
biosynthesis of, 7-51
Cortisol and cortisone metabolism, 7-54
Countercurrent system
and urea excretion, 5-36
of renal tubules, 5-20 ff
Counterpulsation, coronary blood flow, 3-154
Counterrolling of eyes, 8-52
Covalent backbone of DNA, 1-158
Covalent bonds in protein, 1-122
^{51}Cr for platelet turnover rate, 4-74
Cranial nerves
autonomic components, 9-57
classification of, 9-31 ff
visceral components, 9-54
Craniosacral system, 9-53
Creatine phosphate
cardiac muscle contraction, 3-40
in muscle, 1-96
Creatinine and renal dysfunction, 5-38
Creatinine clearance, 5-8, 39
Cremasteric reflex, 9-165
Crenation of erythrocytes, 4-31
Cretinism, 7-41
CRF (corticotrophin releasing factor), 7-20
Critical frequency of fusion, 8-110
Critical temperature, 9-131
Cross-bridges in muscle, 1-76
in sliding filament theory 1-72
Crossed extensor reflex, 9-73
Crossmatching, for blood transfusion, 4-53
Crossover of genes, 1-162
Crush syndrome and renal failure, 5-43
CSF (see Cerebrospinal fluid)
Cupula, vestibular, 9-42
Current, electrical
as stimulus to nerve, 1-38
during voltage clamping, 1-47
in activation of muscle, 1-89
Cushing effect, 3-179
Cushing's syndrome, 7-64
Cutaneous touch sensation, 8-147
Cyanide as poison, 6-78
Cyanide causes hypoxia, 6-63
Cyanocobalamine (see Vitamin B_{12})
Cyanosis, 6-78 ff
definition, 6-64
from defects in heart, 6-76
in bronchial asthma, 6-72
in emphysema, 6-74
in pneumonia, 6-69
Cycle of granulocyte turnover, 4-68
Cycle of lymphocytopoiesis, 4-93
Cycles
female reproductive, 7-101
photochemical, 8-82
Cyclic AMP
activates lipases, 7-129
and insulin secretion, 7-80
and phosphorylase system, 7-119
as purine derivative, 1-142
as second messenger, 7-5
hepatic, effect of glucagon, 7-85
in action of TSH on thyroid cells, 7-36
in catabolite repression, 1-190
in control of protein synthesis, 7-150
in PTH action on bone, 7-45
in secretion of calcitonin, 7-47
increases gluconeogenesis, 7-139
intermediary in action of catecholamines, 7-73
listing of actions, 7-6
mediates ACTH action on adrenal, 7-60
mediates vasopressin activity, 7-27
membrane receptor sites for, 1-6
role in glucogen synthesis, 7-117
stimulates hormone secretion, 7-21
Cylindruria, 5-42
Cystic duct, 2-72
Cystinuria, 1-140
Cytochrome oxidase, 6-78
Cytochrome system in muscle, 1-102
Cytosine, 1-142

Damping of neural circuit, 9-102
Dark adaptation
and receptive fields, 8-89
mechanisms of, 8-100
Dead space
anatomical vs physiological, 6-20
measurement, 6-20, 21
Deafness
categories of, 8-38
causes of, 8-33 ff
diagnosis of, 8-35 ff
treatment of, 8-34
Deamination, oxidative, 1-132

Decarboxylation of amino acids, 1 132
Decerebrate animal, 9-69
Decerebrate rigidity, 9-91
Decerebrate rigidity in man, 9-166
Decerebration, effect on respiratory centers, 6-42
effect on stretch reflexes, 2 72
Decibel notation for sound, 8-11
Decompression chamber and hypoxia, 6 65
Decompression sickness, 6-68
prevention by oxygen, 6-82
Decortication, 9 94
Decortication in man, 9 166
Defecation, 2-119
Defects, optical, 8 68
Deficits, anemias, 4-31 ff
Deficits, endocrine
diabetes mellitus, 7-87
goiter, 7 37
hyper-, hypothyroidism, 7-40 ff
hypophysectomy, protein metabolism, 7-147
of adrenal medulla, 7-75
of adrenocortical function, 7 63
of calcium homeostasis, 7-48 ff
of neurohypophysis, 7-28
releasing factors in therapy, 7-20
Deficits, genetic
aminoacidopathies, 1-135
anemia, 4-20
chronic granulomatous disease, 4 72
congenital adrenal hyperplasia, 7 66
disaccharide intolerance, 2-88
DPG deficiency (see also G6PD), 4 44
errors of metabolism, 1-136
erythroblastosis fetalis, 4-51
G6PD deficiency, 4-29
hemoglobinurias, 4-59
hemophilia, 4-76
of blood clotting, 4-79 ff
of purine, pyrimidine metabolism, 1 153
porphyrias, 4-47
proteinopathies, 1 135
sickle cell anemia, 4-37
spherocytosis, 4-36
thalassemia, 4-36
Deficits, metabolic
genetic basis, 1-136
proteinopathies, 1-135
Deficits, neural, 9-4
hypotheses concerning recovery, 9-6
recovery in cerebral cortex, 9-88
significance of level, 9-161
signs vs symptoms, 9 4
Deficits, neural, clinical, 9-161 ff
Bell's palsy, 9-37
facial paralysis, 9-36
of esophageal motility, 2-102
of swallowing, 2-97
of trigeminal pathways, 9-35
Parkinson's disease, 9-101
spastic hemiplegia, 9-116
spinal cord syndromes, 9-162
Deficits, neural, experimental, 9-5
after cerebellar operations, 9-97
after cortical lesions, 9-113
after lobectomy, lobotomy, 9-118
cerebral decortication, 9-94
cerveau and *encephale isole,* 9-139
from orbitofrontal ablation, 9-117
Deficits, neural, sensory
color blindness, 8-114
in visual system, 8-78
of hearing, 8-33 ff
of olfaction, 8-141
of vestibular function, 8-62
spinal cord lesions, 9-25
Degenerate nature of genetic code, 1 176
Degeneration of nerve fibers, 1-31
Deglutition (see also swallowing), 2-95
Degradation of protein, 7-148
Dehydration of body, 4-116
Dehydrogenase activity, 1-133
Deiodinase, thyroid, 7-33
Deiodination of thyroid hormones, 7-34
Deiter's nucleus, 9-45
Delay, synaptic, central, 9-75
Delayed response, 9-155
Deletion causes mutation, 1-185
Denaturation of protein, 1-127
Dendrites, 1-30
Denervation of muscle, 7 147
Densitometry, microspectral, 8-82
Deoxycorticosterone action in kidney, 5 10
Deoxyribonuclease, 1 151
Deoxyribonucleic acid (see DNA)
Deoxyribonucleotide, 1-146
Deoxyribose in DNA, 1 142
Depolarization of nerve fiber, 1-38
Deprivation, visual, 8-95
Depth perception, 8-117
Dermatomes, 9-23
Dermographism, 3-215
Desoxycorticosterone action in kidney (see Deoxycorticosterone action in kidney) 5-10
Detection, thermal, hypothalamic, 9-135
Detectors (see also Baroreceptors, Chemoreceptors, Mechanoreceptors, etc)
Detectors, neural, 9-1
Detrusor muscle of bladder, 5-49, 55
Deuterium oxide, absorption of, 2 84
Developmental physiology of vision, 8-95
Dexamethasone, 7-56
Dextran filtered in kidney, 5-5
Dextran for plasma volume, 4 119
Diabetes insipidus
and renal dysfunction, 5-47
and thirst, 2-7
from hypothalamic lesion, 9-167
pathogenesis, 7-28
Diabetes insipidus, nephrogenic, 5-46
Diabetes mellitus
clinical features, 7-87
control of ketogenesis in, 7-140
effect on glucose utilization, 7-122
lipid metabolism in, 7 131
Diaphragm, movement of, 6-28
Diarrhea, 2-119
dehydration from, 4-116
loss of bicarbonate ion, 4-109
with disaccharide intolerance, 2-88
Diathermy heating of hypothalamus, 9 134
Dichromats, 8-115
Dicumarol, clinical use of, 4-82
Dielectric constant of membrane, 1-15
Diencephalon, 9-39
components of, 9-20
Diet changes body water, 4-115
Differential thresholds, auditory, 8 31
Differentiation
affected by thyroid hormone, 7-38
in hemopoiesis, 4-15
of fast vs slow muscle, 1-111
of sexual anatomy and function, 7-88 ff
Diffusing capacity for CO, 6-26
Diffusing capacity for O_2, 6-24 ff
Diffusion, 1-17
calculation of rate, 1-19
facilitated, of glucose, 7-115
non-ionic, in kidney, 5 37
of carbon dioxide, 6 27
of gases, theory of, 6-3
Diffusion potentials, 1-33
Digital properties, neural, 9-13
Digitalis, effect on cardiac output, 3-119
Dimer of thymine, 1-168
Dinitrophenol action in kidney, 5 15
Dinitrophenol vs thyroid hormone, 7-39
Diodrast
in renal function tests, 5-39
secretion by tubules, 5 13
use in measurement of blood f, 3 160
Diaphasic action potential, 1-41
Dipolar nature of amino acids, 1 119
Direct inhibition
in spinal reflexes, 9-79
origin and significance, 1 55
Disaccharidases, 2-88
Disaccharide intolerance, 2-88
Discrimination (see name of discrimination)
Discrimination learning, 9 154
Discrimination of frequency, 8-15
Discrimination threshold for frequency, 8-32
Disease (see name of disease, or organ or system affected)
Disparity, retinal, 8-120
Disseminated (multiple) sclerosis, 9 170
Dissociation curves (see also oxyhemoglobin dissociation curves)
carbon dioxide, 6-16
combined, CO_2 and O_2, 6-71
oxygen, 6-11 ff
Dissociation of amino acids, 1-120
Disulfide bridges
in denaturation of ribonuclease, 1 128
in insulin molecule, 7-76
Diuresis, 5 18 ff
and urea excretion, 5 35
osmotic, 5-25
Diuretics, 5-26
Diuretics, mercurial, 5-15
Dizziness, 8-59
DNA, (see also Nucleic acid)
discovery of, 1-158
role in cardiac hypertrophy, 3-45 ff
DNA and genes, 1 136
DNA and mutations, 1-138
DNA ligase, 1-166
DNA molecules
genetic organization of, 1 162
single stranded, 1-166
size and shape, 1-161
structure of, 1-158
DNA of nucleolus, 1-6
DNA polymerase, 1-166
DNA replication, 1-163
and corticosteroid actions, 7-57
and granulocytopoiesis, 4-66
and lymphocytopoiesis, 4-93
folic acid requirement for, 4 38
in cycle of hemopoiesis, 4-13
in erythropoiesis, 4-22
regulated by steroids, 7-7
stimulated by androgen, 7-99
DNA replication time, 4-14
DNA structure, 1-142
DNA vs. RNA, 1-142
Dominance in genes, 1-136
Dominant hemisphere
lateralization of function, 9-120
localization of speech mechanisms, 9-148
Donath-Landsteiner hemolysin, 4-59

Donnan ratio and CO_2 transport, 6 18
Dopa and dopamine biosynthesis, 7-71
Dopa for Parkinson's disease, 9-101
Dopamine
and gonadotropin secretion, 7-15
as transmitter, 1 59
distribution in CNS, 9-83
in basal ganglia, 9-101
Dopamine-β-oxidase, 9 62
Dorsal root potentials, 9-11
Dorsal roots, visceral afferents in, 9-54
Dorsal spinocerebellar tract, 9-25
Dorsomedial thalamic nuclei, 9-39
Double helix, structure of, 1-60
Downhill vs uphill transport, 1-22
DPG (2,3-diphospho D-glyceric acid)
binding to hemoglobin, 4-44
concentration in stored blood, 4 55
Drinker respirator, 6-84
Drinking
control of, 9-125
rhythm of, 9-103
Drug effects
cause edema, 4-89
emetics, 3-110
in epilepsy, 9-172
in motion sickness, 8-61
on bone marrow, 4-19
on capillary permeability, 4-87
on colonic motility, 2-119
on gallbladder, 2 77
on gastric secretion, 2-45
on granulocytopoiesis, 4-71
on H^+ concentration, 4 104
on liver function, 2-67
on secretion of saliva, 2-29
on taste, 8-131
Drugs excreted by kidney, 5-37
Dubin-Johnson syndrome, 2-67
Ducts, salivary, 2 17
Ductus arteriosus
failure to close, 6-77
fetal circulation, 3-238 ff
Ductus venosus, age of closure of, 3 239
Duodenal loops, 2-113
Duodenal motility, 2 111
Duodenal rhythmicity, 2 113
Duodenal ulcer, 2-45
Duration of visual perception, 8-112
Dwarfism
pituitary, 7-143
thyroidal, 7-41
Dyads of retina, 8-75
Dynamic state of body constituents, 4-7
Dynamic vs static receptors, 8-44
Dysarthria, 9-150
Dysbarism, 6-68
Dysphagia, 2-103
Dyspnea, 6-59 ff
from defects in heart, 6-76
miscellaneous causes, 6-61

Ear and hearing, 8-12 ff
Early receptor potential, 8-86
Edema, 4-88 ff
cardiac, with renal dysfunction, 5-44
caused by aldosterone, 5-47
in renal dysfunction, 5-43 ff
nephrotic, 5-42
pulmonary, with dyspnea, 6-60
Edema fluid, 4-84
Edinger-Westphal nucleus, 9-32
EDTA and Ca binding, 7-44
EDTA-like relaxing factor, 1-94
EDTA prevents clotting, 4-76
EEG (see electroencephalogram)
Effective renal plasma flow, 5-15
Effectors, neural, 9-1
Efferent auditory pathways, 8-28
Efferent cochlear bundle, 9-37
Efferent nerve fibers, 9-74
in olfactory system, 8-138
in visceral system, 9 54
in visual system, 8-80
to inner ear, 8-25
to vestibule, 8-47
Efficiency of ventilation, 6 37
Einthoven method of vector analysis, 3-66
Eisenmenger's complex, 6 77
Elastic element
parallel, in muscle, 1 65
series, in muscle, 1-81
Elastic tissue of lung, 6-75
Elasticity of muscle, 1 79
Electrical properties
of cell membranes, 1-15
of living systems, 9-12
of muscle fiber membrane, 1 92
Electrical recording, 9 9 ff
Electrical stimulation, 9 8 ff
during conditioning, 9-152
in study of sensation, 8-5
of olfactory receptors, 8 135
of otolith system, 8 50
of salivary secretion, 2-27
of taste receptors, 8 130
self-stimulation of brain, 9-154
Electrical transmission, 1 59
Electrical vs chemical transmission, 9-82
Electrocardiogram, 3-64 ff
abnormalities of, 3-69 ff
digitalis effect on, 3 73
effect of myocardial injury on, 3-62
electrodes for standard leads, 3-67
left bundle branch block effect on, 3 72
myocardial ischemia effect on, 3 72
procaine amide effect on, 3-73
quinidine effect on, 3-73
right bundle branch block effect on, 3-69
time intervals during cardiac cycle, 3-104
vector analysis, 3 65 ff
ventricular hypertrophy effect on, 3 69
Electrocardiography, 3-58 ff
Electrochemical potential
in active transport, 1-21
in carrier transport, 1-21
Electroconvulsive shock and memory, 9-158
Electrode potentials, 1-33
Electrodes for electromyography, 1-106
Electrodermal response to hearing, 8-36
Electroencephalogram (EEG), 9-106 ff
activation of, 9-137
of sleep, 9-137
Electrolyte concentrations
in bile, 2-60
in blood plasma, 4-2
in gastric juice, 2-36
in lymph, 4-85
in pancreatic juice, 2-50
in perilymph, endolymph, 8-18
in saliva, 2-22
in succus entericus, 2-80
Electrolyte depletion
in Addison's disease, 7-63
in adrenal hyperplasia, 7-68
with water depletion, 4-116
Electrolytes
affect aldosterone secretion, 7-63
and membrane potentials, 1-35
intestinal absorption, 2-85
pancreatic secretion of, 2 53
taste of, 8-121
transported in gallbladder, 2-74
Electromagnetic radiation spectrum, 8-64
Electromyography, 1-104 ff
Electron microscopy of membranes, 1-3
Electron microscopy of muscle, 1-63 ff
Electron transport
in HCl secretion, 2 40
in mitochondria, 1-11
in nerve, 1-51
in phosphorylation in muscle, 1 102
Electrophoresis of plasma proteins, 4-4
Electrophysiology of reflexes, 9-74
Electrophysiology of sensation, 8-3
Electroretinogram (ERG) 8-86
Electrostatic bond hypothesis of muscle, 1 78
Electrotonic potentials (see also potentials, electrotonic), 9-11 ff
Elongation of polypeptides, 1 179
Embden-Meyerhof pathway in RBC, 4-28
Embolus following thrombosis, 4-80
Embryology
of nervous system, 9 20
of sarcoplasmic reticulum, 1-70
sexual differentiation, 7-92
Emetics, 2-110
Emotion, 9-143 ff
and gastric secretion, 2-47
Emphysema, 6-73 ff
changes in lung volumes, 6 72
detection of uneven ventilation, 6-21
dyspnea in, 6-59
volume-flow relations, 6-34
Encephale isole, 9-139
End plate of muscle, 1-57
End plate potentials
and acetylcholine release, 1 57
and quanta of acetylcholine, 1-89
as computational process, 9 13
Endocochlear potential, 8 22
Endocrine glands, 7-1 ff
neuroendocrine relations, 7 2
Endocytosis as transport, 1-6
Endocytosis in thyroid cells, 7-33
Endogenous sources of amino acids, 1-130
Endolymph, 8-40
electrolytes of, 8-18
Endometrium, cycles of, 7-102
Endonuclease, 1-151
Endoplasmic reticulum (ER), 1-7
Energy balance, 2-1
hypothalamic control of, 9-127
Energy exchange of nerve, 1-51
Energy generation in muscle, 1-96 ff
Energy of radiation, 8-64
Energy release in muscle, 1-104
Energy reserves in adipose tissue, 7-123
Energy requirements for vision, 8 84 ff
Energy sources for muscle, 7-112
Energy sources in man, 7-125
Energy transmission in hearing, 8-15
Enterogastric reflex, 2-10
and gastric emptying, 2-109

Enterogastrone
and gastric emptying, 2-109
and gastric secretion, 2-41
inhibits gastric secretion, 2-44
Enterohepatic circulation
of bile pigments, 2-68
of bile salts, 2-63
Enterokinase, 2-51
Environment, rotation of, 8-54
Enzyme action
in energy generation in muscle, 1-98
in fibrinolysis, 4-77
in membrane transport, 1 4
in mitochondrial membrane, 1-12
in phagocytosis, 4-63
in protein degradation, 7-149
Enzyme control, types of, 7-114
Enzyme deficit in gout, 1-155
Enzyme deficit in orotic aciduria, 1-156
Enzyme induction and repression, 1-187
Enzyme induction by steroids, 1-193
Enzyme pathways
for blood clotting, 4-75 ff
for glycogen synthesis, 7-117
for lipid synthesis, 7-124
for protein synthesis, 1-174
for purine synthesis, 1-147
for pyrimidine biosynthesis, 1-145
for RNA biosynthesis, 1-170
for synthesis of catecholamines, 7-71
in amino acid metabolism, 1-132
in control of lipogenesis, 7 141
Enzyme reactions and odor perception, 8-143
Enzyme recognition site, 1-174
Enzyme secretion by pancreas, 2-54
Enzymes (see also name of enzyme)
disaccharidases, 2-88
in gastric juice, 2-37
in pancreatic juice, 2-51
in saliva, 2 24 ff
in succus entericus, 2-82
Eosinophils of blood, 4 62
effect of cortisol on, 7-58
Ephapse, 1-59
Epilepsy
clinical description, 9-170 ff
effect on EEG, 9-108
irritation of area 4, 9-115
Epilepsy, psychomotor, 9-147
Epinephrine, 7-71
activation of phosphorylase, 7-118
and regulation of body temperature, 9-136
as autonomic transmitter, 9-62 ff
as insulin antagonist, 7-86
causes paralytic secretion of saliva, 2-26
conjugation and inactivation, 9-63
effect on cardiac output, 3-118
effect on gluconeogenesis, 7-137
effect on heart muscle, 3-116
in pheochromocyte, 7-73
increase in coronary blood flow, 3-138
released by acetylcholine, 9-61
renal blood flow effects of, 3-236
Equal-loudness contours, 8-32
Equivalent electrical circuit
for EPSP, 1-54
for muscle membrane, 1-91
of nerve fiber, 1-39
Endoplasmic reticulum (ER), rough and smooth, 1-8
Errors, translational, protein synthesis, 1-177
Erythroblastosis fetalis, 4-51
Erythrocyte defects, 4-34
Erythrocyte destruction, 4-34 ff
Erythrocyte dimensions, 4-32
Erythrocyte life span, 4-30
hemoglobin turnover, 4-36
Erythrocyte membrane
in hemolysis, 4-57
pores in, 1-17
water permeability, 1-27
Erythrocyte precursors, 4-24
Erythrocyte sedimentation rate, 4-60
Erythrocytes
cell indices, calculation, 4-31
deformability of, 4-61
destroyed in spleen, 4-35
formation in bone marrow, 4-11
hemoglobin as buffer, 4 105
in blood gas transport, 6-13 ff
intracorpuscular defects, 4-36
lifespan in stored blood, 4-54
metabolic pathways, 4-27
structure and function, 4-26 ff
variations in, 4 30
Erythron, defined, 4-25
Erythropoiesis, 4-21 ff
effect of cortisol, 7-58
from common stem cell, 4-18
regulation of, 4 25
steady state model, 4 26
suppression by hemoglobin, 4 23
Erythropoietic factor, renal, 4-26
Erythropoietin, 4-25
Esophageal sphincter
lower, 2-102
upper, 2-98
Esophagogastric junction, 2-102
Esophagus
and swallowing, 2 95 ff
disturbances of motility, 2-102
motility of, 2-100
pain from, 2-104
spasm of, 2-103
stricture of, 2-103
Essential amino acids, 1-129
Esterification in lipid synthesis, 7-127
Estrogen
adrenal synthesis of, 7-51
effect on LH secretion, 7-106
in reproductive cycle, 7-102
mechanism of action, 7-107
model of cellular action, 7-7
transport and metabolism, 7-106
Estrogen secretion
by human placenta, 7-110
in male, 7-101
in phases of ovarian cycle, 7-105
Ethyl alcohol
and vestibular function, 8-60
effect on gastric secretion, 2-46
Ethyl iodide, arteriovenous, measurement of cardiac output, 3-159
Eukaryotic cells, chromosomes of, 1-193
Eustachian tube and deafness, 8-34
Evans blue for plasma volume, 4-118
Evaporation of water for heat loss, 9-130
Evoked potentials, cortical, 9-105
Evolution of active transport, 1-24
Evolution of blood, etc., 4-1
Evolution of membranes, 1-16
Evolution of plasma proteins, 4-4
Exchange transfusion, 4-51
Excitation-contraction coupling, 1-95 ff
Excitation in retina, 8-89
Excitation of muscle, 1-61 ff
Excitation of nerve, 1-43
Excitatory postsynaptic potential (EPSP)
and chloride ions, 9-78
in spinal reflexes, 9-76 ff
origin and significance, 1-52 ff
Exercise, hyperpnea during, 6-57
Exocrine glands, control of, 9-53 ff
Exocytosis
as transport process, 1 6
in anterior pituitary cells, 7-11
Exogenous sources of amino acids, 1-129
Expeditions, high altitude, 6-65
Expiration
control of, 6 41 ff
movements of chest, 6-28 ff
pressure changes, 6 34 ff
pressure volume relations, 6-37
volume changes, 6-30 ff
Expiratory center
location, 6-43
recording activity from, 6-45
Expiratory reserve volume, 6-31
Extensor reflexes, decerebrate, 9-91
Exteroceptors, 8-1
Extracellular fluid, 4-113
Extracellular vs intracellular H^+, 4-110
Extraocular muscles, 8-75
Extrinsic factor in pernicious anemia, 4-37
Eye (see also Visual system), 8 70
anatomy of, 8-70
optical properties, 8-66 ff
Eye color, genetics of, 1 136
Eye movement, 8-96
from vestibular stimulation, 8-51

F-actin, 1-65
Facial nerve, autonomic outflow, 9-57
Facial nerve (VII), 9 35 ff
taste fibers in, 8-126
Facial paralysis, 9 36
Facilitated diffusion of glucose, 7 114
Facilitation, neural, 9 4
in spinal reflexes, 9 76
Factors, official names
in blood clotting, 4-75
Facultative water reabsorption, 5-19
Fallot, tetralogy of, causes hypoxia, 6-77
Fasting
control of ketogenesis in, 7-140
lipid metabolism in, 7-131
water retention in, 4-116
Fat, absorption of, 2-91
Fat excreted in feces, 2-57
Fat tissue, circulation at rest vs maximum dilatation, 3 166
Fatigue in cochlea, 8 21
Fatty acid CoA, 7-128
Fatty acid metabolism, 7 123 ff
Fatty acid oxidation cycle, 1 132
Fatty acid synthesis, 7-123
Fatty acids
affect initiation of peptide chains, 7-153
as insulin antagonists, 7-86
hepatic ketogenesis from, 7-140
in membrane structure, 1-16
oxidation inhibits glycolysis, 7-120
sources for lipid synthesis, 7 128
stimulate gluconeogenesis, 7 138
Fava beans and anemia, 4-29
Fecal fat excretion, 2-57
Feces, composition of, 2-120
Feces, defecation of, 2-119
Feedback control
of ACTH secretion, 7-61
of anterior pituitary cells, 7-22
of cochlear mechanics, 8-13
of granulocytopoiesis, 4-70
of hypothalamic secretion, 9-124

of neurohumoral systems, 7 14
of pyrimidine biosynthesis, 1-145
of thyroid gland, 7-36
vestibular, for movement, 8-40
Feedback loop, 9-18
Feeding (see also Food intake)
Feeding center, hypothalamic, 2-2
Feeding reflexes, 2-3
Fenestration for deafness, 8 34
Fenn effect in muscle, 1-86
Ferrihemoglobin (see Methemoglobin), 4-46
Ferritin, 4 39
Fetal circulation, defects related to hypoxia, 6-76
Fetal hemoglobin dissociation curve, 6-12
Fetoplacental unit, 7-109
Fetuin, 4-6
Fetus, blood flow in, 3-238 ff
Fever as set point change, 9-136
FFA of plasma, 7-123
Fibrin in blood clotting, 4-77 if
Fibrinogen, 4-75
as coagulant, 4-82
conversion to fibrin, 4-77
variations in plasma levels, 4-7
Fibrinolysis, 4-77
Fick method for measurement of blood flow, 3-158 ff
Fick principle for renal blood flow, 5-15
Filaments of muscle, 1-62
Filling of stomach, 2-107
Filtration, glomerular, 5-4 ff
filtration fraction, 5-15
measurement of rate, 5-4 ff
of protein, 5-41
First breath of newborn infant, 6-40
Fissures of cerebellum, 9-43
Flagella and cilia, 1-13
Flavoprotein enzymes in muscle, 1 102
Flexion reflex, 9-70 ff
Flicker fusion frequency, 8-110
Flocculonodular lobe
anatomy of, 9-46
functions of, 9-98
Fluoride ion in saliva, 2-23
Fluoroscopy of stomach, 2-107
Flux, transmembrane, 1-18
Flux, unidirectional vs net, 1-25
Folates prevent anemia, 4-38
Folding protein theory, 1-72
Folia, cerebellar, 9-45
Follicle, ovarian, cycle of, 7-101
Follicle-stimulating hormone (see also FSH)
action on ovary, 7-102
control of testis tubules, 7-99
Food as source of water, 4-114
Food intake
control of, 2-1 ff
factor in energy balance, 9-127
rhythmicity of control of, 9-103
Foot, cutaneous blood flow in, 3-214
Foramen ovale, patent
causes hypoxia, 6-77
Force in muscle contraction, 1-73
Force-velocity relationship in muscle, 1-84
Form discrimination, 8-108
Formylation of amino acids, 1-179
Fovea of retina, 8-73
Fragility of RBC, 4-58
Free water clearance, 5-22
Freeze-etch technique, 1-4
Freezing of stored blood, 4-54
Frequency coding, auditory, 8-27
Frequency discrimination, 8-15
Frequency discrimination threshold, 8 32
Frequency modulation, neural, 9 103
Frequency of nerve impulses, sensory, 8 7
Frequency of sounds, 8-11
Frequency response of ear, 8-32
Frequency selectivity, 8-28
Frohlich's syndrome, 2-2
Frontal eye fields, 9-117
Frontal lobe syndrome, 9-118
FSH secretion
control by steroid levels, 7-105
Functional residual capacity of lung, 6-31
Fundic glands of stomach, 2 32
Fundus of stomach, motility of, 2-104
Fusion of cells, 1-193

γ-Aminobutyric acid, 9-82
as transmitter, summary of, 1-59
γ Efferent system, 9 84
γ-Globulin, 4-7
γ Motor fibers, 9-83
in decerebrate state, 9-93
G-actin, properties of, 1-65
G6PD deficiency, in erythrocytes, 4 29
Galactose absorption, 2-86
Gallbladder
anatomy of 2-72
contractions of, 2-73
emptying, 2 76
functions of, 2-73
Gallstones, 2-77 ff
Galvani, Luigi, 9-9
Ganglia, autonomic, 9-55 ff
Ganglion (see also name of ganglion)
Ganglion cells in retina, 8-74
Gas (see name of specific gas)
Gas bubble, P-V relation, 6-39
Gas in intestine, 2-120
at high altitude, 6-67
Gases, techniques for measuring, 6-7 ff
Gases, theory and laws, 6-1 ff
Gases, transport by blood (see Carbon dioxide transport, Oxygen transport)
Gastrectomy for ulcer, 2 46
Gastric (see also Stomach)
Gastric distension
and food intake, 2 4
and water intake, 2-7
Gastric emptying, 2-108
Gastric filling, 2-107
Gastric fistula and food intake, 2-1
Gastric juice, 2-35 ff
Gastric motility, 2-104 ff
abnormalities of, 2-109
types of contractions, 2 105
Gastric pacemaker, 2-106
Gastric peristalsis, 2-106
Gastric secretion, 2-31 ff
control of, 2-41 ff
enzyme content, 2-37
HCl and electrolytes, 2 36
measurement of acidity, 2-47
normal course of, 2-44
phases of, 2-41 ff
Gastric secretory capacity, 2-45
Gastric ulcer and HCl secretion, 2-45
Gastrin, 2-15
effect on gastric secretion, 2-42 ff
in control of secretion, 2-41
Gastrocnemius, white muscle, 1-109
Gastrocolic reflex, 2-10
Gastroduodenal pressure cycle, 2-107
Gastroileal reflex, 2-116
Gastrointestinal (see also names of viscera, guts, etc)
Gastrointestinal gas at high altitude, 6 67
Gastrointestinal tract, circulation at rest vs maximum vasodilatation, 3 166
Gastrone, 2-38
Gay-Lussac's law of gases, 6-2
Gender recognition, 7-88
Gender role, 7-94
Gene activity, 1-187
and nucleic acids, 1 158
regulation of, 1-186
Gene expression, 1-193
Generation of neural signals, 9-16
Generator potential, sensory, 8 6
Genetic code, 1-175 ff
Genetic organization of DNA, 1-162
Genetic recombination, 1-162
Genetic sex, 7-88
Genetics of eye color, 1-137
Geniculate bodies (see Lateral geniculate body)
Geniculostriate bundle, 8-78
Genital ducts, differentiation, 7 92
Genital dystrophy, hypothalamic, 9-167
Genital sex differentiation, 7-92
Genitalia, differentiation of, 7-93
Genotype governs controls, 1-136
Genotype vs phenotype, 1-137
Geometric horopter, 8-120
Giant edema, 4-5
Gigantism, 7-143
Glands (see also Endocrine glands, Exocrine glands, or name of gland)
salivary, 2 17
Glands of small intestine, 2-79
Glands of stomach, 2-31
Glia cells
functions of, 1-31
varieties of, 9-16 ff
Globin
bilirubin binding, 2-65
structure of, 4-42
Globin variants, 4-45
Globulins, immunoglobulins, 4 101
Globus pallidus, 9-50
effect of lesions, 9-100
signs of lesion, 9-161
Glomeruli, renal
filtration in, 5-4 ff
filtration of Na^+, 5-24
filtration of protein, 5 41
measurement of filtration, 5-6 ff
structure, 5-1
Glomerulonephritis, 5-40
edema in, 5-44
Glomerulotubular balance, 5-12
Glossopharyngeal nerve (IX), 9-37
autonomic outflow, 9-57
taste fibers in, 8-126
to salivary glands, 2-20
Glucagon
as gastrointestinal hormone, 2 15
biosynthesis of, 7-78
chemical nature, 7-78
control of secretion of, 7-76 ff
Glucagon action
in hepatic cells, 7-85
increases cyclic AMP, 7-80
increases gluconeogenesis, 7-136
Glucocorticosteroids (see also Cortisol, Corticosteroids)
as insulin antagonists, 7-86
biosynthesis of, 7-51
control of synthesis, 7-60
increase lipolysis, 7-130
protein catabolic effect, 7-135

Gluconeogenesis
in adipose tissue, 7-127
irreversible steps, 7-132
pathways and control, 7-132
regulated by glucagon, 7-85
Gluconeogenesis vs lipogenesis, 7-140
Glucoreceptors, hypothalamic, 2-5
Glucose
intestinal absorption, 2-86
precursor for lipid synthesis, 7-123
stimulates insulin release, 7 79
Glucose-activation unit, 7 112
Glucose-6-phosphate dehydrogenase (see G6PD)
Glucose tolerance test
influence of growth hormone, 7-146
oral vs intravenous, 7-81
Glucose transport
by facilitated diffusion, 7-114
by renal tubules, 5-9 ff
inhibited by growth hormone, 7-145
passive, in muscle, 7-114
Glucose utilization
by erythrocytes, 4-29
by heart muscle, 7-121
energy generation in muscle, 1-97
energy yield of reactions, 1-104
Glucostatic hypothesis, 2-5
Glucuronyl transferase, 2-66
Glutamic acid as NH_2 donor, 1 134
Glutamine and gout, 1-156
Glutamine as source of NH_3, 5-31
Gluteal reflex, 9 165
Glycerinated muscle, 1-93
Glycerol in fat absorption, 2-92
Glyceroneogenesis, 7-127
Glycine in bile salts, 2 62
Glycogen homeostatic unit, 7-112
Glycogen metabolism in muscle, 1 97
Glycogen phosphorylase, 7-117
Glycogen synthesis, 7-117
Glycogenic amino acids, 1-132
Glycogenolysis
and energy generation, 1-98
regulated by glucagon, 7-85
Glycolysis in erythrocytes, 4-29
Glycolysis in muscle, 1-97
control of, 7-112
details of enzymatic processes, 1-99 ff
irreversible steps, 7-132
Glycolysis in nerve, 1 50
Glycolytic unit, 7-112
Glycoproteins of plasma, 4-4
Glycosuria
in acromegaly, 7 144
renal, 5-46
Goiter, 7-37
Goldman equation, 1-36
Golgi apparatus
and pancreatic secretion, 2 55
of anterior pituitary cells, 7-11
of neurons, 1-29
of thyroid gland, 7-29
structure and function, 1-10
Golgi tendon organs, 9-84
Gonadal sex differentiation, 7-91
Gonadotrophin release, cycles of, 7-94
Gonadotrophin releasing factor, 7 19
Gonadotrophs of anterior pituitary gland, 7-9
Gout, 1-153
Graded potential, 8-6
Gradient, intestinal, 2-112
Grand mal epilepsy
clinical description, 9-171
EEG pattern, 9-108
Graft rejection, 4-98
Graft vs. host reaction, 4-19
Granulocyte turnover rate, 4-66
Granulocytes, 4-62 ff
circulating in blood, 4 67
Granulocytic leukemia, 4 71
Granulocytopoiesis, 4-64 ff
from common stem cell, 4 18
hazards of chemotherapy, 4 71
Granulomatous disease of childhood, 4 72
Grasp reflex, decorticate, 9 94
Grasping movements, 9-116
Graves' disease, 7-40
Gravity on otoliths, 8 49
Gravity, zero
adaptation to, 8-54
test for vestibular function, 8 49
GRF (Growth hormone releasing factor), 7-19
Growth
abnormalities of, 7 143
affected by thyroid hormone, 7-38
cellular processes of, 7-146
hormonal control of, 7-143 ff
stimulated by androgen, 7 99
Growth hormone (GH)
amino acid sequence, 7 144
as insulin antagonist, 7-86
chemical nature, 7-143
localization in cells, 7 12
standard curve, assay of, 7 4
Growth hormone action, 7-143
increases glucose output from liver, 7-135
increases lipolysis, 7 130
like insulin action, 7-144
mobilizes lipid, 7 146
on glucose transport, 7 145
Growth hormone releasing factor (GRF), 7 19
Growth rate in hemopoiesis, 4-15
Guanine, 1 142
Guanylic acid, 1 149
Gustatory nucleus, 9 38

Habituation, vestibular, 8-54 ff
Hair cells
auditory, 8-21
for touch, 8-150
of saccule, 8 46
of utricle, 8-44
vestibular, 8 42
Haldane method
gas mixtures analysis, 6 7
sampling of alveolar air, 6-4
Hand, cutaneous blood flow in, 3-214
Haptoglobins, 4 6
Hashimoto's disease, 7 41
Head, Sir Henry, and aphasias, 9 149
Hearing, 8-10 ff
cortical localization of, 9-120
Hearing aids, 8-39
Hearing and speech, 8-29
Hearing, binaural, 8-28
Hearing deficits, 8 38
Heart (see also Myocardial, Myocardium)
absolute refractory period, 5-53
action potential, 3 52, 3-93
adrenal medullary catecholamines in regulation of, 3-115
anatomy, 3-25
anxiety effect on volume output, 3-118
aortic coarctation effects on circulation of, 3-138
arrhythmias, 3-75 ff
atrial contraction, 3-104
atrial fibrillation, 3 81 ff
atrial flutter, 3-80 ff
atrial systole, 3-103
atrioventricular block, 3-85
atrioventricular node, 3-56
autoregulation, rate and rhythm effects on, 3-110 ff
baroreceptor reflexes and venoconstriction, 3 180, 181
blood circulation of, 3-130 ff
blood supply to, 3-131 ff
blood volume during contracture, 3-98
bypass procedures for blood flow, 3-154, 155
calcium ion in excitation-contraction of, 3-36 ff
cardiac cycle sequence, 3-101 ff
cardiac massage, 3-153
chemoreceptor reflexes, 3 181 182
circulation in anemic states 3 138
conductance velocity 3 54
conduction system 3 54
contraction force, 3-87, 88
coronary artery collateral circulation, 3-143 ff
development of collateral circulation to, 3 143
diastasis, 3 103
dilatation, 3-43
effective refractory period, 3 53
electrophysiology, 3-48 ff
embryonic differentiation of myocytes of, 3 35
equilibrium in left and right chambers 3 124 ff
excitation contraction coupling, 3 36 ff
exercise effects on circulation of, 3 141, 142
extrasystoles, 3 77 ff
failure, 3-140
glucagon effect on regulation of 3 116
hypertension effects on circulation of, 3 141
hypertrophy, 3 41
hypothermic effects on circulation of, 3-140
impulse conduction after injury, 3 62
inferior cardiac nerve, 3 113
initiation of conductance impulse, 3-57
innervation of, 3-112, 113
internodal conduction of impulse, 3 55
isovolumetric contraction, 3 101
isovolumetric relaxation, 3 103
left ventricle reflexes, 3-185
maximum ejection, 3 101
measurement of output and work, 3 116 ff
mechanical activity, 3 86 ff
middle cardiac nerve, 3 113
muscle T tubules, compared with skeletal, 3 32
myocardial bundles of, 3 27
myocardial capillary circulation, 3 130
myocardial cells of, 3 28
myocardial tension, 3 110
myofilaments, 3-29 ff
nerves of, 3 113
nervous stimulation of blood flow in, 3-139
oxygen metabolism, 3-130
pacemaker cells, 3-28
pacemaker potential, 3-53
paroxysmal tachycardia, 3 79
percentage distribution of blood from, 3-2
phasic irregularity, 3-85
posture effect on volume output, 3-118
premature contractions, 3-77
protodiastole, 3-102
pulsus alternans, 3-84
pump action of, 3-98 ff
rapid filling, 3-103

reduced ejection, 3-102
reduced filling, 3 103
reflex effect on vasomotor tone, 3 184
regulation of, 3 110 ff
relative refractory period, 3-53
resting potential, 3-52
RNA in hypertrophy of, 3 46
sarcoplasmic reticulum, 3-31
sinoatrial block, 3-85
sinoatrial node, 3 55
sinus arrythmia, 3-84
sinus bradycardia, 3-85
sounds, 3-107 ff
spread of excitation impulse, 3-57, 58
stimulation of collateral circulation to, 3-145 ff
stress state response of, 3-136, 137
superior cardiac nerve 3-113
tachycardia, 3 137
temperature effect on autoregulation, 3 112
threshold potential, 3-50
transfusion effects on circulation of, 3 137
transmembrane potential, 3 48 ff
vagus nerve regulation of, 3-114
valves, movement during cardiac cycle 3-105 ff
valvular defect effects on circulation, 3-140
venoconstriction, 3-180, 181
ventricular fibrillation, 3-83
volume output, 3-118 ff
weight, percentage of body weight, 3-26
work of, and glucose utilization, 7-121
work of, effect on glucose entry, 7 116
work, power, and efficiency, 3 96 ff
workload in hypertrophy, 3 46

Heart defects, cause hypoxia, 6 77

Heart rate
aortic and atrial pressure, 3 11
at high altitude, 6-67
blood loss effect on, 3-186
cardiac reserve in, 3-129
carotid sinus pressure, 3-10
chemoreflex control of, 3-185
comparison of, during rest and exercise, 3-117
determination of steady state, 3-9
exercise and excitement effects on, 3-119 ff
nervous control of, 3-112
relation to carotid sinus pressure, 3-9
stress state effect on, 3-136
ventricular output control, 3 11

Heart ribosomes, control of aggregation of, 7-154

Heartburn, 2-104

Heat edema, 4 89

Heat loss, control of, 9-129

Heat of recovery in muscle, 1-87

Heat production
and temperature regulation, 9-131
by nerve fibers, 1-51
in skeletal muscle, 1-85

Heavy meromyosin (HMM), 1-67

Heavy water (see Deuterium oxide), 2-84

Helium in O_2 therapy, 6-82

Helix of myosin, 1-67

Hematurias, 5-40 ff

Heme-heme interaction, 4-43

Heme synthesis, 4-41

Heme synthetase, 4-42

Hemiplegia, levels of, 9-162

Hemiplegia, spastic, 9-116

Hemisection of spinal cord, 9-25

Hemochromatosis, 4-40

Hemocytometer, 4-32

Hemodynamics, 3 22

Hemoglobin, 4-41 ff
abnormal, 6-12
as buffer for H^+, 4-105
binding to organic phosphates, 4 44
binding with oxygen, 4-43
changes at high altitude, 6-66
component chains, 4 42
effect on erythropoiesis 4-23
fetal, 6 12
filtered in glomeruli, 5 4
hemes and O_2 transport, 6 17
in cyanosis, 6-79
isohydric shift, 6 17
oxygen capacity of, 6-6

Hemoglobin and bile pigments, 2-65

Hemoglobin degradation, 4-35

Hemoglobin derivatives, 4 46

Hemoglobin variants, 4-36 ff
and globin variants, 4-45
component chains of, 4 42
in sickle cell anemia, 4 37

Hemoglobinopathies, 4-45

Hemoglobinurias, 4-59

Hemolysis, 4 57
and jaundice, 2 67
and suspension stability, 4-57 ff
in stored blood, 4 55
in vivo, 4-59
normal and abnormal, 4-34

Hemolysis vs hemopoiesis, 4-36

Hemolytic anemia and G6PD, 4 29

Hemolytic disease of newborn, 4-51

Hemophilia, 4-79
deficiency of blood clotting, 4-76

Hemopoiesis, 4-10 ff

Hemopoiesis vs hemolysis, 4-36

Hemorrhage
causes vasopressin release, 7-26
hemostasis following, 4 73
physiological changes following, 4 120
water and electrolyte movements, 4-117

Hemostasis, 4 73 ff

Henderson-Hasselbalch equation
and CO_2 transport, 6-6
and H^+ in urine, 5-28
and plasma buffers, 4-104
calculations from, 4 112
for amino acid buffers, 1-118
nomogram for, 4-111

Heparin, 4-77
clinical use, 4-82

Hepatolenticular degeneration, 9-168
ceruloplasmin deficiency, 4-6

Hereditary orotic aciduria, 1-156

Hering-Breuer reflex, 6-47 ff
in pneumonia, 6-72

Hernia, hiatal, 2-103

Heteroplastic hemopoiesis, 4-13

Heterozygous state, 1-136

Hexokinase deficiency in RBC, 4-44

Hexokinase in muscle, 7-117

Hiatus hernia, 2-103

Hierarchy of neural controls, 9-15

High altitude
acclimatization, 6-66
and erythropoiesis, 4-26
hypoxia, 6-65

High pressure breathing, 6-83

Hippuric acid excretion, 5-13

Hiroshima nuclear bomb, 4-62

His, bundle of, 3-56

Histamine
and inflammation, 4-64
effect on cardiac output, 3-118
effect on gastric secretion, 2-45
relation to asthma, 6-73

Histocompatibility antigens, 4-100

Histones as regulators, 1-193

Hodgkin-Huxley equivalent circuit, 1-36

Hodgkin-Huxley theory of action potential, 1-48 ff

Hoffmann's reflex, 9-87

Holger-Nielsen resuscitation, 6 83

Holoenzyme of polymerase, 1-170

Homeostasis, 9-58
of body fluids, 4-1
of calcium, 7-42 ff
of triglycerides, 7 126

Homoiothermic animals, 9-128

Homoplastic hemopoiesis, 4-12

Homozygous state, 1-136

Hopping reactions, 9-114

Horizontal cells of retina, 8 75

Hormone secretion, 7-21

Hormone-sensitive lipase, 7 129

Hormone synthesis, 7-22

Hormones, 7-1
assay of concentrations, 7 3
control lipid metabolism, 7-130
control lipolysis, 7 130
gastrointestinal, 2 15 ff
general mechanisms of action, 7-5
localization by immunohistochemical methods, 7-10
synthesis vs release, 7 3

Hormones in lactation, 7-109

Hormones of pregnancy, 7-107

Horopter, geometric, 8-120

Horsley-Clarke technique, 9 5

Human chorionic gonadotropin (HCG), 7-108

Human voice in communication, 8-29

Human vs animal sensory studies, 8-4

Hunger, 2-1 ff
and insulin secretion, 7-84

Hunger contractions, 2-4
as type II waves, 2-105

Huntington's chorea, 9-170

Hydrochloric acid
in gastric juice, 2 36
measurement of, 2-47
mechanism of secretion, 2-38
secretion by parietal cells, 2 34

Hydrogen bond in base pairing, 1-151

Hydrogen bonds in DNA, 1-124

Hydrogen bonds in protein, 1 123

Hydrogen ion activity
and multiple factor theory, 6-56
disturbances of regulation, 4-106 ff
effect on Hgb-O_2 binding, 4-43
effect on respiration, 6 56
in extracellular fluids, 4-103
in fluids of body, 4-104

Hydrogen ions
and acidification of urine, 5-27
and signal generation, 9-16
buffer systems, 4-104
chemodetection, central, 6-47
excretion and fixation of, 4-104
form NH_4^+ in kidney, 5-31
metabolic sources of, 4-103
secreted in stomach, 2-39
sensitivity of central detectors, 6-53

Hydrogen sulfide and hemoglobin, 4-47

Hydrolysis of triglyceride, 7-127

Hydrophilic groups in membrane, 1-15

Hydrophilicity and diffusion, 1-19

Hydrophobic bonds in protein, 1-125

Hydrophobic groups in membrane, 1-15

Hydrostatic pressure in kidney, 5-6

Hydroxy-amino acids, 1-133

17-Hydroxycorticosteroids, 7-55
1-Hydrozinophthalazine, renal blood flow effects of, 3-236
Hyperactivity, hypothalamic, 9-142
Hyperaldosteronism, 7 70
Hyperalgesia, 8-156
Hyperbaric oxygen chamber, 6-22
Hypercalcemia and calcitonin, 7-48
Hypercapnia in asphyxia, 6-64
Hypercortisolism, 7-66
Hypermetropia, 8-69
Hyperphagia, hypothalamic, 2-2
and rage responses, 9-144
Hyperplasia as growth process, 7-146
Hyperpnea during exercise, 6-57
Hyperpolarization, neural, 1-38
in retinal units, 8-88
Hypersensitivity, delayed, 4-98
Hypersexuality, 9-146
Hypertension
coronary blood flow in, 3-141
in adrenal hyperplasia, 7-69
in glomerulonephritis, 5-40
in nephrosclerosis, 5-41
Hyperthyroidism, 7-40
Hypertonic fluids, 4-122
injection of, 4-116
Hypertrophy as growth process, 7-146
Hyperuricemia in gout, 1-153
Hyperventilation
causes alkalosis, 4-109
effect on alveolar air, 6-5
Hypnotoxin, 9-142
Hypochromic anemias, 4-32
Hypogammaglobulinemia, 4-102
Hypoglossal nerve, 9-38
Hypoglycemia
effect on food intake, 2-4
effect on gastric secretion, 2-46
stimulates GH secretion, 7-143
stimulus for glucagon release, 7-80
Hypophosphatemia, renal, 5-46
Hypophysectomized animal, 7-60
Hypophysectomy, 7-147
medical, 7-23
Hypophysis (see Pituitary gland)
Hypoproteinemia and edema, 5-44
Hypoprothrombinemia, 4-80
Hypothalamic island, 9-126
Hypothalamic lesions, 9-167
Hypothalamic-pituitary-adrenal axis, 7-61 ff
Hypothalamus, 9-40 ff
and sham rage, 9-144
as neuroendocrine transducer, 9-124
autonomic connections, 9-65
control of anterior pituitary gland, 7-13 ff
control of energy balance, 9-127
control of food intake, 2-2
control of Leydig cells, 7-95
control of pituitary-thyroid axis, 7-36
control of vasoconstriction, 3-168
control of water balance, 9-126
control of water intake, 2-6
effect on testis tubules, 7-99
Frohich's syndrome, 2-2
negative feedback of sex steroids on, 7-105
neuroendocrine relations, 7-2
nuclear groups, 9-41
pathways and connections, 9-42
regulation of body temperature, 9-131
restraining mechanism, 9-144
sex differentiation of, 7-93
suppressor mechanism, 9-141
thermal detection, 9-135
thermal set point, 9-136
Hypothyroidism, 7-41
effect on numbers of cells, 7 146
Hypotonic fluids, 4-122
Hypoxia, 6-63 ff
and cyanosis, 6-80
and DPG levels in RBC, 4-45
and erythropoiesis, 4-25
causes in pneumonia, 6 70
effect on EEG 9-108
summary of causes, 6-27
Hypoxia, cardiac, and glucose utilization, 7-121
Hysteresis in ventilation, 6-38

Idiopathic epilepsy, 9-171
IgG (γ **globulin**), 4-101
IgM (macroglobulin), 4-101
Ileal rhythmicity, 2-113
Ileocecal valve, 2-116
Ileus, paralytic, 2-115
Illusion, oculogyral, 8-53
Immune response, 4-99
in cell transplants, 4-19
role of lymphocytes, 4-90
theories of, 4-91
Immunity, cell mediated, 4-101
Immunoglobulins, 4-101
primary and secondary response, 4 90
Immunohistochemistry, 7-10
Immunological deficits, 4-102
Immunological paralysis, 4-92
Immunosuppression, 4-100
Impedance of muscle membrane, 1-92
Impedance of tympanic membrane, 8-12
Inactive vs active state of nerve, 1-51
Incompatability, tissue, 4-19
Incomplete dominance of genes, 1-137
Index of mitosis, 4-14
Inducer vs. repressor, 1-187
Induction of enzymes, 1-187
Infants
anemia of, 4-32
blood flow in, 3-238
development of urinary continence, 5-55
immunological deficits, 4-101
resuscitation, 6-85
Infection and hemolysis, 4-60
Inferior colliculus, 8-26
Inflammation
and edema, 4-88
effect of cortisol, 7-59
granulocytes in, 4-64
Information transmission
in auditory system, 8-29
in sensory systems, 8-1
non-neural, 9-16
via nerve impulses, 9-12
Infrared absorption and odor, 8-143
Infrared technique for CO_2, 6-8
Inheritance
of ABO blood groups, 4-50
of anosmia, 8-141
of color blindness, 8-117
Inhibition, neural, 9-2
from sensory stimuli, 8-8
in cerebellum, 9-49
in retina, 8-89
mechanisms of, types, 1-55
of stretch reflex, 9-73
types of, mechanisms, 9-79
Inhibition, noncompetitive, 1-20
Inhibition of protein synthesis, 1-181
Inhibition vs. repression, 1-145
Inhibitors of RNA synthesis, 1-171
Inhibitory mechanism
from cerebral cortex, 9-145
in auditory system, 8-28
in cochlea, 8-24
Inhibitory postsynaptic potential (IPSP), 1 54 ff
in spinal reflexes, 9-77
Initiation of peptide chains, 1 179
hormonal control of, 7 153
Injury, protection from, 8-144
Innervation
of digestive tract, 2-8 ff
of esophagus, 2-96
of gastric galnds, 2-35
of pancreas, 2 49
of salivary glands, 2-19 ff
of taste receptors, 8-123
of vestibular system, 8 46
Inosinic acid, 1-147
Input-output analysis, 9-4
Insensible perspiration, 4-115
Insensible water loss, 4-114
Insertion causes mutation, 1-185
Inspiration
control of, 6-41 ff
movements of chest, 6 28 ff
pressure changes, 6 34 ff
pressure-volume relations, 6 37
volume changes, 6-30 ff
Inspiratory capacity of lung, 6 31
Inspiratory center, 6-43
recording activity from, 6-45
Inspiratory reserve volume, 6-31
Instructive theory of antibodies, 4-91
Insulin
A- and B-chains, 7-77
chemical structure of, 7-76
comparative physiology of, 7 76
history of discovery of, 7-76
rate of destruction, 7-82
Insulin action
and glucose entry in muscle, 7 116
antilipolytic, 7-130
evokes hunger, 7-84
inhibits gluconeogenesis, 7-115
inhibits protein degradation, 7-148
on cyclic AMP in fat cells, 7 129
on gastric secretion, 2 46
on initiation of peptide chains, 7-153
on lipoprotein lipase, 7 128
Insulin antagonists, 7-85
Insulin binding
on membranes, 7-81
to antibodies, 7-86
Insulin biosynthesis, 7-77
Insulin deficiency, 7-87
Insulin-glutathione transhydrogenase, 7-82
Insulin molecule, 1-184
Insulin secretion, 7-76 ff
Insulin turnover, 7-81
Integrated control
of CHO metabolism, 7-122
of lipid metabolism, 7-130
Integrative action, 9-4
central computation in, 9-13
for emotional expression, 9-144
in regulations, 9-121
in temperature regulation, 9-136
Intelligence and frontal lobes, 9-118
Intensity of sounds, 8 11
Interauricular septal defect, causes hypoxia, 6-77
Interconnections of basal ganglia, 9-50
Intercostal muscles, 6-29
Intermodality effects, visual, 8-80
Internal capsule
anatomy of, 9-49
components of, 9-51 ff
signs of lesion, 9-161
Internuncial neurons, 9-2
Interoceptors, 8-1

Interpleural pressures, 6-38
Interpleural (intrapleural) space, 6-32
Interstitial cell stimulating hormone (see LH)
Interstitial fluid, 4-85
Interventricular septal defect, causes hypoxia, 6-76
Intestinal absorption, 2-83 ff
of carbohydrate, 2-86
of electrolytes, 2 85
of fat, 2-91
of protein and amino acids, 2-89
of vitamins, 2-93
of water, 2 84
transamination during, 2-90
Intestinal gas, 2-120
Intestinal obstruction, 2-115
dehydration from, 4-116
Intestinal pacemaker, 2-112
Intestinal villi, 2-114
Intestine, large, motility, 2-117
Intestine, law of, 2-114
Intestine, small
disturbances of motility, 2-115
motility of, 2-111 ff
peristalsis, 2-113
polar conduction, 2-115
site of PTH action, 7-46
Intestino-intestinal reflex, 2-10
Intracellular fluid, 4-113
Intracellular vs extracellular H^+, 4-110
Intracodon compensation, 1-185
Intrathoracic (interpleural) pressure, 6-38
Intrinsic factor in pernicious anemia, 4-37
Intrinsic muscles of eye, 8-71
Intubation of intestine, 2-84
Inulin clearance, renal, 5-6, 7
test for renal function, 5-39
Iodide transport, 7-31
Iodide trapping, 7-30
in non-thyroidal tissues, 7-30
Iodination of tyrosine, 7-31
Iodine
deficiency of intake, 7-41
in thyroid hormones, 7-29
metabolic pathways, 7-31
plasma levels of, 7-33
protein-bound forms, 7-34
Iodine, radioactive
in insulin assay, 7-76
in radioimmunoassay, 7-4
tests of thyroid function, 7-39
Iodotyrosine in thyroid, 7-32
Ion distribution, cellular, 1-35
Ionic currents in muscle, 1-92
Ionic flux in EPSP, 1-53
Ionic pumps and action potential, 1-49
Ionizing radiation
effect on hemopoiesis, 4-10
effect on immune response, 4-100
effect on lymphocytes, 4-97
suppression of bone marrow, 4-71
Iron
daily requirement, 4-39
deposits in thalassemia, 4-36
in heme synthesis, 4-42
in hemoglobin-oxygen binding, 4-43
intestinal absorption, 2-85
radioactive, 4-26
transport in plasma, 4-6
Iron deficiency anemia, 4-32
Iron metabolism, 4-39
Isoantibodies and blood groups, 4-49
Isocortex, cerebral, 9-104
Isoelectric point of amino acid, 1-119
Isohydric shift of hemoglobin, 6-17
Isometric contraction of muscle, 1-82
Isoproterenol and β receptors, 9-64
Isotonic contraction of muscle, 1-79
Isotonic fluids, 4 122
Itching, 8-157

Jackson, J Hughlings, 9 15
Jacksonian epilepsy, 9-171
Jaundice
hemolytic, 2-67
nonhemolytic, 2-67
Jaw jerk, 9 87
Jejunal iron absorption, 4-39
Jejunal rhythmicity, 2-113
Jugular vein, venous pulse pattern, 3-199
Junction, esophagogastric, 2 102
Junction potentials, 1-34
Junctional complex in cells, 1-14
Junctions, electrical, 1-60
Junctions, tight, 1-14
Juxtaglomerular apparatus, 5-25
Juxtamedullary glomeruli, 5-1, 2

Kallikrein in saliva, 2-24
Kallikrein-kinin-kininase system, 4 64
Karyotype, lymphocyte
female, 7-89
male, 7-90
Kent, accessory bundle of, 3-57
Ketoacidosis, 7-87
Ketogenesis, control of, 7-139
Ketogenic amino acids, 1 132
Ketosis, 7-140
Ketosteroids and androgens, 7-96
17-Ketosteroids in urine, 7-55
Kidney (see also Glomeruli, Tubules, and specific processes), 5-1 ff
blood flow, 3-228 ff
cell separation theory of blood flow in, 3-233
central blood volume effects on circulation in, 3-235
circulation at rest vs maximum dilatation, 3-166
cortex vs medullary blood flow, 3-231
exercise effect on blood flow in, 3-235, 236
hypothermic effects on blood flow, 3-237
intrarenal pressure theory of blood flow in, 3-233
measurement of blood flow in, 3-160
methods for measurement of blood flow in, 3-228
myogenic control of blood pressure in, 3-232
nervous control of blood flow in, 3-234
oxygen consumption and blood flow, 3-237, 238
postural effects on blood flow in, 3-235, 236
sodium feedback theory of blood flow in, 3-233
Kidney disease, 5-40 ff
edema in, 4-88
Kidney function, 5-1 ff
abnormalities of, 5-38 ff
acid-base balance, 5-27 ff
effect of aldosterone, 7-59
excretion of water, 5-17 ff
filtration fraction, 5-16
measurement of blood flow, 5-15 ff
regulation of $[H^+]$, 4-106
tests for abnormalities, 5-38 ff
urea excretion, 5-34 ff
Kidney secretes erythropoietin, 4-26
Kidney structure, 5-1
Kinase, pancreatic, 2-51
Kinetics of hemopoiesis, 4-13 ff
erythropoiesis, 4-22
granulocytopoiesis, 4-65
lymphocytopoiesis, 4-97
Kinocilium
auditory, 8-19
vestibular, 8-42
Kluver-Bucy syndrome, 9-145
Knee jerk as reflex, 9-86
Korotkoff method of auscultation, 3-150
Krypton, coronary blood flow measurement, 3-161
Kupffer cells, 2-60

Labor, oxytocin in, 7 28
Labyrinthine reflexes, 9-95
in decerebrate state, 9-91
Labyrinthine systems, 8-40 ff
Labyrinthotomy, 8-61
Lactate production
and gluconeogenesis, 7 134
effect on $[H^+]$, 4-106
in anaerobic muscle, 7-113
Lactogen, placental, 7-109
Lactose operon, 1-188
Landolt ring test, 8 107
Landsteiner's law of blood groups, 4-49
LaPlace's law and bladder function, 5-52
Large intestine, nerve supply, 2-9
Larynx in swallowing, 2-98
Late receptor potential, 8-86
Latency relaxation of muscle, 1-82
Latent heat of blood, 4-2
Latent period of muscle, 1-79
Lateral geniculate body
color coding in, 8-91
layers of, 8-77
Lateral inhibition in retina, 8-88
Lateral nuclei of thalamus, 9-39
Lateralization
in auditory system, 8-26
in cerebral cortex, 9-119
Law of intestine, 2-114
Learning, 9-151
and discriminations, 9-154
and memory, 9-157
as recovery process, 9-7
stages of, 9-158
Lecithin in bile, 2-69
Length-tension relation in muscle, 1-82
Lengthening reaction, 9 92
Lens
aging of, 8-68
properties of, 8-72
Lentiform nucleus, 9-50
Lesch-Nyhan syndrome, 1-155
Lesions, neural (see also Deficits, neural), 9-4
Leukemia
chemotherapy of, 4-19
granulocytic, 4 71
lymphocytic, 4-10
Philadelphia chromosome, 4-19
Leukocytosis inducing factor, 4-67
action on marrow, 4-71
Levels of function, auditory, 8-27
Levels of neural control, 9-15
Leydig cells, 7-95
LH (Lutemizing Hormone)
action on ovary, 7-102
control of Leydig cells, 7-95
in plasma, 7-16
LH releasing factor, 7-95
LH secretion
control by steroid levels, 7-105
feedback control loops, 7-14
in phases of ovarian cycle 7-105

Lieberkuhn, crypts of, 2-79
Life span of cells, 4-18
erythrocytes, 4-30
erythrocytes, stored, 4-54
lymphocytes, 4-93
Light, 8-63 ff
Light adaptation, 8-104
Light meromyosin (LMM), 1-67
Lightning pains in tabes, 9-166
Limbic system, 9-146 ff
and orbitofrontal cerebral cortex 9-117
autonomic connections, 9 65
Linear period of isotonic contraction, 1-84
Lingual papillae, 8-123
Lipase
gastric, 2-38
pancreatic, 2-52
Lipase and gluconeogenesis, 7-138
Lipase clearing factor, 7-128
Lipases in lipolysis, 7-129
Lipid (see also Fat, Fatty acids, Cholesterol)
Lipid-lipid interactions, in membranes, 1-1
Lipid metabolism
and corticosteroid actions, 7 57
control of, 7-130
Lipid solubility, 1-19
Lipids in lymph, 4-85
Lipogenesis, control of, 7-140
Lipogenesis vs gluconeogenesis, 7-140
Lipolysis
affected by glucagon, 7 85
regulation of, 7-129
Lipoprotein lipase, 7-128
Lipoproteins of plasma, 4-4
Liver, 2-60 ff
actions of glucagon, 7-85
affected by drugs, 2 67
bile secretion, 2-60 ff
blood flow, 3 226
circulation at rest vs maximum dilatation, 3-166
conjugation of thyroid hormones, 7-34
control of gluconeogenesis in, 7-133
control of ketogenesis in, 7-139
in hemoglobin degradation, 4 35
source of blood glucose, 7-83
vascular resistance in, 3-226
Liver cell components, 1-172
Load, renal parameter, 5-9
Lobar pneumonia, 6-69
Lobe (see name of lobe)
Local contraction of muscle, 1 71
Local excitatory state (LES)
as excitation process, 9-13
in stimulation of nerve, 1 39
Localization, neural
of cerebral motor control, 9-112
of pain stimulus, 8-155
of sound stimulus, 8-33
of speech, cortical, 9-120
of touch stimulus, 8-151
Locomotor ataxia, 9-165
Loewi, Otto, 9-59
Long-acting thyroid stimulator (LATS), 7-40
Longitudinal fasciculus (medial), 9-29
Loudness contours, equal, 8-32
Loudness of sound, 8-33
Loven reflex, 3-180
Lower motor neurons, 9-161
LTH secretion, control of, 7-15
Luminance thresholds, 8-103
Lung
alveolar air, 6-4
alveolar gas exchange, 6-13
blood pressure and flow regulation, 3-204
compliance, 6-31
exercise effects on circulation in, 3 208
fetal blood flow in, 2 239
gas diffusion in, 6-4 ff
gas exchange schema, 6-22
hysteresis in pressure-volume, 6-39
in pneumonia, 6-70
mechanoreceptors, 6-48
phasic blood flow, 3-206
physical model of alveoli, 6-39
pressure pulse in circulation, 3-206
pressure-volume relation, 6-32
pulmonary vessel reflexes, 3-184
respiratory effects on blood pressure and flow, 3 207
sheet flow concept of circulation in, 3 207
surface activity, 6-38 ff
vascular blood flow resistance in, 3-205
vascular waterfall concept of circulation, 3 207
vasomotor control of circulation in, 3 209
Lung volumes, 6-30
changes in asthma, 6 72
changes in emphysema, 6-74
changes in pneumonia, 6 71
effect of pressures upon, 6-31
measurement of subdivisions 6 31
Lupus erythematosus of kidney, 5-40
Lutein, ovarian, 7-102
Luteinizing hormone (see LH)
Luteotropic hormone (LTH), 7-15
Lymph, 4-3
and fat absorption, 2-92
composition of, 4-85 ff
from salivary glands, 2-20
lymphocyte content, 4-95
production and flow rates, 4 86
Lymph nodes
and lymph flow, 4 84
migration of lymphocytes, 4 97
Lymphatic, cannulation of, 4 97
Lymphatic obstruction and edema, 4-89
Lymphatic system, 4-83 ff
Lymphocyte flow, 4 99
Lymphocyte karyotypes, 7-89 ff
Lymphocyte pool, 4-97
Lymphocyte transforming factor, 4-98
Lymphocytes, 4-92 ff
circulation and migration, 4-96
role in immune response, 4-100
Lymphocytic leukemia, 4-10
Lymphocytopoiesis, 4-93
Lymphoreticular system, 4-90 ff
Lymphotoxin, 4-99
Lysosomes
as cell organelles, 1-10
degrade protein, 7-148
of granulocytes, 4-62
Lysozyme
in gastric juice, 2-38
in saliva, 2-24

M, N and P antigens, 4-51
Macrocytic anemias, 4-33
Macromolecules and pinocytosis, 1-25
Macula adherens, 1-14
Magenstrasse, 2-107
Magnesium ions
and ATPase activity of myosin, 1-93
and phosphorylase system, 7-119
Malate cycle in lipogenesis, 7-126
Malate in gluconeogenesis, 7-133
Male estrogen secretion, 7-101
Malnutrition causes edema, 4 89
Maltose, absorption of, 2-88
Mammary glands, 7 109 ff
Mammary myoepithelial cells, 7 28
March hemoglobinuria, 4-59
Marrow transplantation, 4 18
and lymphocytopoiesis, 4-95
Masking of sounds, 8-32
Mastication, 2-95
Maturity-onset diabetes, 7-87
Maximal tension in muscle, 1-83
Maze learning, 9-155
Mechanoreceptors (see also Baroreceptors)
for touch, 8 148
in lung, 6 48
medial forebrain bundle, 9-42
Medial longitudinal fasciculus, 9-29
vestibular contribution to, 8 49
Median eminence, hypothalamic, 7-13
Medical hypophysectomy, 7-23
Medulla (see also Adrenal medulla)
Medulla oblongata, 9 30 ff
autonomic centers, 9 65
centers for regulations, 9-125
neurons sensitive to CO_2, 6-46
respiratory centers, 6 42 ff
sleep, activating system, 9 140
vomiting center, 2 110
Megakaryocytopoiesis, 4-74
Megaloblastic anemia, 4-37
Meissner's corpuscle, 8 146
Melanin in substantia nigra, 9 101
Melanocyte stimulating hormone (see MSH)
Membrane binding in DNA replication, 1-166
Membrane, capillary
permeability of, 4-85
variations in permeability, 4-87
water, electrolyte exchange, 4 122
Membrane, cell
composition of, 1-1
electrical properties, 1 15
of erythrocyte, 4-27
permeability to urea, 5 35
structure of, 1-15
unit of, 1-3
Membrane models, 1-16
Membrane of mitochondrion, 1-11
Membrane of nucleus, 1 7
Membrane permeability
and hormone secretion, 7-21
and water movement, 1 28
changes during nerve impulse, 1-49
hormonal control, 7-5
mathematics of, 1-18
to steroid hormones, 7-5
Membrane potential (see also Potentials, electrical, Potentials, electrotonic), 1-34 ff
gastric glands, 2-36
in salivary glands, 2-28
Membrane receptor sites
for female steroid hormones, 7-107
for insulin, 7-82
Membrane transport
and membrane structure, 1-4
mathematics of, 1-17 ff
of amino acids, 1-139
of glucose in muscle, 7-112
Membranes
cochlear, 8-14 ff
phylogenetics of, 1-26
Memory, 9-157 ff
after frontal lobectomy, 9-118
chemical correlates, 9-159
stages of, 9-158
Memory trace, 9-157
Menière's disease, 8-61

Menstrual cycle, 7 101
Mercurial diuretics, 5-15
Merkel's disc, 8-147
Meromyosins, 1-67, 3 33
cross-bridges of, 1-73
Mesencephalic root of nerve V, 9-35
Mesencephalon, 9-21
structure of, 9-31
Mesentary, blood flow in, 3-224
Messenger RNA, 1 172
and immune response, 4 91
and nucleolar function, 1-6
in control of protein synthesis, 7-149
in granulocytopoiesis, 4-69
in synthesis of plasma proteins, 4 7
regulated by steroids, 7 7
translation of, 1 179
turnover of, 7-149
Metabolic inhibitors in nerve, 1-50
Metabolic rate
affects lymph flow, 4-88
and thyroid function, 7-38
daily turnover of substance, 6 70
dyspnea when increased, 6-61
effect of temperature, 9 133
effect upon respiratory exchange, 6-19
Metabolic regulation, 7-112
Metabolism in muscle, 1-112
Metabolism of amino acids, 1-131 ff
Metamorphosis and thyroxin, 7 38
Metencephalon, 9-21
structure of, 9-31
Methemoglobin
concentration in blood, 4-46
ferric iron in, 4-43
relation to cyanide poisoning, 6-78
3-O-methyl glucose transport, 7-115
Methylene blue for HCN poisoning, 6 78
Micelle model of membrane, 1 16
Micelles in fat absorption, 2 91
Micelles of bile salts, 2-63
Microcirculation of RBC, 4-61
Microcytic anemias, 4-32
Microelectrode techniques, 9 10
Microelectrodes (see also Single unit analysis)
for stimulation of muscle, 1-71
in study of salivary glands, 2-28
Microfilaments, 1-14
Microglia, 1-31
Microphonic potentials, cochlear, 8-20
Micropuncture technique
and water excretion, 5 19
for glucose transport, 5-10, 11
for study of kidney, 5-3
Microsomes, 1-8
Microspectral densitometry, 8-82
Microtubules, 1-13
Micturition, 5 54, 55
Midbrain, autonomic outflow, 9-57
Middle ear mechanisms, 8-13
Midline nuclei of thalamus, 9-39
Migration inhibitory factor, 4-98
Milk ejection or let-down, 7-111
Milk ejection reflex, 7-27
Milk synthesis, 7-109
Mineralocorticoid (see Aldosterone)
Miniature end plate potentials, 9-13
Minimum detectable acuity, 8-106
Minimum energy for vision, 8-85
Minimum separable acuity, 8-106
Minute volume, respiratory, 6-52 ff
Mitochondria, 1-11
active transport of ions, 1-24
and thyroid hormone action, 7-39
electron micrograph of, 1-9
enzyme reactions in muscle cells, 1-103
functional states of, 1-12
in gluconeogenesis, 7 132
in lipid synthesis, 7-124
of muscle cells, 1-65
of nerve fiber, 1-29
Mitochondrial membranes, 1-104
Mitoses in erythropoiesis, 4-22
Mitoses in granulocytopoiesis, 4 66
Mitoses, index of, 4-14
Model of DNA, 1-161
Models of base pairs, 1-152
Modern theory of renal function, 5 3
Molecular disease, 1 136
Molecular size and diffusion, 1-19
Monkey, cerebral cortical operations, 9-114
Monoamine oxidase, 7 72
in autonomic system, 9 63
Monochromats, 8-115
Monocular depth perception, 8-118
Monoglyceride shunt, 2-92
Monoglycerides, absorption, 2 91
Monomers of cells, 1 1
Monosynaptic reflex, 9-1
direct inhibition of, 1 55
inhibition of, 9 80
properties and significance, 9-75 ff
Morphine on gut motility, 2-119
Mossy fibers of cerebellum, 9-45
Motility of gut
colonic, 2 117
colonic, abnormalities, 2-120
esophageal, 2-100
esophageal, disturbances, 2-102
gastric, 2-104 ff
gastric, abnormalities, 2 109
intestinal, 2-111 ff
intestinal, disturbances, 2 115
Motion sickness, 2 110
and cerebellar nodule, 9 98
and vestibular system, 8 59
Motivation and food intake, 2 3
Motor cortex, precentral, 9-89
Motor end plate, 9-67
Motor mechanisms, cortical, 9-112 ff
Motor nerve endings, 1-111
Motor nerves to muscle, 1-89
Motor nerves, types, 9-74
Motor neuron pool, 9 71
Motor neurons, location, 9-66
Motor pathways in CNS, 9-27 ff
Motor unit, 1-80
numbers of fibers, 1-105
Mountain climbing expeditions, 6-65
Mouth, movements, 2-95
Mouth-to-mouth resuscitation, 6-83
Movement, 9-96 ff
Movement and lymph flow, 4-88
Movement, associated, automatic, 9-100
Movement from muscle contraction, 1-108
Movement, voluntary
and basal ganglia, 9-100
and cerebellum, 9-97
and vestibular function, 8-40
as rhythmic phenomenon, 9-103
control of, 9-96 ff
Movements of eye, 8-75
types, and control, 8-96
Moving-fork mechanism, for DNA, 1-163
MSH molecule, 7-60
MSH vs ACTH, 7-61
Mucus in bile, 2-73
Mucus in colon, 2-82
Mucus in saliva, 2-18
Mucus in stomach, 2-38
Mullerian duct, differentiation, 7-92
Multiple allelism, 1-137
Multiple factor theory, 6 56
Multiple myeloma, 4 101
Multiple sclerosis, 9-170
Multiplier effect of releasing factors, 7-18
Multisynaptic reflexes, 9 75
Muscle, cardiac, glucose utilization by, 7-121
Muscle, cardiac vs skeletal, ribosome cycle, 7-155
Muscle cells, ions of, 1 35
Muscle, electron microscopy of, 1-63 ff
Muscle, mechanical properties, 1-79 ff
Muscle, red vs white, 9-90
comparative physiology of, 1 109
Muscle, skeletal, 1-61 ff
aggregation of ribosomes, 7 154
CHO metabolism, 7-112 ff
control of, 9-66 ff
differentiation of fibers, 1 111
energy generation, 1 96
excitation-contraction coupling, 1 95 ff
Fenn effect, 1-86
gluconeogenesis during exercise, 7 134
growth in adolescence, 7-146
in control of blood glucose, 7 83
models of electrical properties, 1 91
organization of function, 1-106
proteins after denervation, 7-147
speed vs histology, 1-109
theories of contraction, 1 72 ff
tone of, 9-89
twitch vs slow fibers, 1 110
work-induced growth, 7 147
Muscle, skeletal vs cardiac, carbohydrate metabolism of, 7-122
Muscle, smooth
control of, 9-53 ff
intrinsic, of eye, 8-67 ff
Muscle, smooth vs skeletal, 1-83
Muscle spindles, 9-82 ff
Muscle, tibialis anticus, 9-71
Muscles, extraocular, 8-75
Muscles of deglutition, 2-96
Muscles of respiration, 6 28 ff
control of activity, 6-49
Muscles of urinary tract, 5-48 ff
Muscularis mucosae, intestinal, 2-114
Mutation
in molecular disease, 1-138
molecular basis of, 1-183
Myelencephalon, 9-21
Myelin, 1-32
Myelin sheath, 1 30
Myelinated fibers, autonomic, 9-55
Myelocytes in granulocytopoiesis, 4-65
Myeloma, multiple, 4-101
Myenteric plexus, 2-8
Myenteric reflex, 2-114
Myocardial cells
automaticity of impulse, 3-54
cardiac hypertrophy, 3-45
conductivity, 3-35
contraction of, 3-35
depolarization in electrocardiography, 3-58
embryonic development of, 3 33, 34
excitability of impulse, 3-54
excitation-contraction coupling, 3-45
regeneration of, 3 43
repolarization in electrocardiography, 3-58

Myocardial infarction, electrocardiogram, 3-73
Myocardial sarcomere, 3-28 ff
A band, 3-30 ff
calcium ion storage and release, 3-36 ff
compared with skeletal muscle, 3 31, 32
desmosomes, 3-31
fascia adherens of, 3-31
in cardiac dilatation, 3-43
macula adherens of, 3-31
mitochondria of, 3-31
Z line, 3-30
Myocardium
actin concentration in, 3-32
anatomy, 3-29
anoxia effects on, 3-131
basal state oxygen consumption, 3-131
circulation at rest vs maximum vasodilatation, 3-166
compensated blood flow in ischemia, 3-145
contractile active state, 3-39
contraction, compared with skeletal muscle, 3 36
contraction force, 3-87, 88
depolarization in electrocardiography, 3-58, 59
energy sources compared with skeletal muscle, 3-42
intercalated discs, 3-31
mechanochemical coupling, 3 42
metabolism of, 3-41, 3-135
myosin concentration in, 3-32
oxygen consumption in, 3-39, 3 135, 136
P cells, 3-28
Purkinje cells, 3-28
repolarization in electrocardiography, 3 60
transitional cells, 3-28
tropomyosin concentration in, 3-32
Myoepithelial cells
mammary, 7-28
salivary, 2-18
Myofibrils, 1-68
Myofilaments, 3-29 ff
Myoglobin and blood flow, 1-112
Myoglobin vs hemoglobin, 4-47
Myopia, 8-68
Myosin, 1-67
arrangement in sarcomere, 3-30
concentration in myocardium, 3-32
interaction with tropomyosin and troponin, 3-38
Myosin filament model, 1-77
Myosin molecules, 1-73
Myotatic reflex, 9-72 ff
and tone of muscle, 9-90
inverse, 9-85
Myotomes, 9-23

Nagasaki nuclear bomb, 4-62
Narcolepsy, 9-167
Narcosis, nitrogen, 6-83
Natural righting, 9-98
Nausea, 2-13 ff
Neck reflexes, 9-95
in decerebrate state, 9-91
Negative control by operon, 1-189
Negative feedback
neuroendocrine, 9-124
of sex steroids on pituitary gland, 7-105
Negative staining of muscle, 1-74
Neocerebellum, 9-47
Neonate, blood flow in, 3-238 ff
Neopallium, 9-104
Nephrogenic diabetes insipidus, 5 46
Nephron, unit of kidney, 5-1
Nephrosclerosis, arteriolar, 5-41
Nephrotic syndrome, 5-40
Nernst equation
for concentration cell, 1-33
for transmembrane potential, 1-35
Nerve (see name of nerve)
Nerve cell (see Neuron), 9-2
Nerve deafness, 8-34
Nerve endings
for touch, 8-145
in cochlea, 8-24
in muscle spindles, 9-85
on twitch vs slow fibers, 1-111
Nerve fibers, 1-37 ff
all or-none principle, 1-43
classification and properties, 1-41
classification of, 9-74
compound action potential, 1-41
degeneration and regeneration, 1-31
diameter vs conduction speed, 1-42
distribution in retina, 8-73
from vestibule, 8-47
general description, 1-30
intermediary metabolism, 1-50
maturity vs conduction speed, 1 42
refractory periods of, 1 45
Nerve impulse, 1-29 ff
comparative physiology of velocity, 1-43
Nerve impulses
from retinal units, 8-89
from sensory stimuli, 8-7
patterning of, 9-12
recording techniques, 9-10
Nerve roots, spinal, 9-23
Nervus intermedius of Wrisberg, 9-35
Net vs unidirectional flux, 1-25
Neural circuits, 9-102
Neural coding, sensory, 8-9
Neural control (see Control, neural)
Neural crest, 9-20
Neural deficits (see Deficits, neural)
Neural pathways
pain, 8-157
touch, 8-151
Neural response, sensory, 8-6
Neural signals, coding, 9-12
Neural tracts of CNS, 9-24 ff
Neural tube, 9-20
Neurilemma, 1-31
Neuroendocrine relations, 7-2
Neuroendocrine transduction, 9-124
Neurofibrils, 1-29
Neuroglia, 9-16 ff
Neurohumoral control
future studies of, 7-22
of anterior pituitary gland, 7-13 ff
Neurohypophysial disorders, 7-28 ff
Neurohypophysis, 7-24 ff
ADH and water excretion, 5-18
chemical structure of hormones, 7-26
embryology of, 7-24
Neurological disorders (see also Deficits, neural), 9-161 ff
Neuroma, 1-32
Neuromuscular transmission, 9-67
Neuron doctrine, 9-68
Neuron types, 9-2, 3
Neuronal conditioning, 9-156
Neurons, 1-29
ions in, 1-35
Neurosecretion
by hypothalamus, 7-13
in posterior pituitary gland, 7-24
Neutropenia, congenital, 4-72
Neutrophils of blood, 4-62
Nissl bodies, 1-29
Nitrites, effect on cardiac output, 3-118
Nitrogen
during oxygen therapy, 6-82
from amino acids, 1-131
in dysbarism, 6-68
partition in blood constituents, 4-8
Nitrogen analyzer, 6-8
Nitrogen balance, 1-134
Nitrogen dilution method, 6-21
Nitrogen narcosis, 6-83
Nitrous oxide
arteriovenous, measurement of cardiac output, 3-159
coronary blood flow measurement, 3-161
Nocturia, 5-42
Nocturnal dyspnea, paroxysmal, 6-60
Nocturnal hemoglobinuria, 4-60
Nodes (see also Lymph nodes)
Nodes of Ranvier, 1-31
Nodule, cerebellar, 9-98
Noeud vital (for respiration), 6-42
Nomogram for H-H equation, 4-111
Nomogram, Siggaard-Andersen, 4-112
Noncompetitive inhibition, 1-20
Nonhemolytic jaundice, 2-67
Non-ionic diffusion, renal, 5-37
Nonprotein nitrogen
and renal dysfunction, 5-38
normal and abnormal levels, 4 8
Nonsense codons, 1-176
Norepinephrine, 1-59
as *Acceleransstuffe,* 9-60
as autonomic transmitter, 9-62
details of biosynthesis, 7-71
distribution in CNS, 9 83
in sleep deprivation, 9-143
increase in coronary blood flow, 3-138
renal blood flow effects of, 3 236
Nuclear sex, 7-88
Nucleated red cell precursors, 4-24
Nuclei of cerebellum, 9-46
Nuclei of muscle cells, 1-65
Nucleic acid metabolism, 1-150
hormonal control of, 7-143
Nucleic acid replication, 1-6
and female sex steroids, 7-107
Nucleic acids (see also DNA, RNA)
and gene activity, 1-158
of plasma, 4-3
turnover of, 1-151
Nucleolus of cells, 1-6
Nucleolus of neuron, 1-30
Nucleosides, 1-142
Nucleotides, 1-142
as phosphorylated nucleosides, 1-142
structure and nomenclature, 1-143
synthetic, 1-175
Nucleus, neural (see name of nucleus)
Nucleus of cells, 1-6 ff
Nucleus of neuron, 1-30
Nystagmus, 8-55
as test for vestibular function, 8 52

3-O-methyl glucose transport, 7-115
Obesity and food intake, 2-2
Obesity, hypothalamic, 9-167
Obligatory water reabsorption, 5-19
Obstruction, intestinal, 2-115
Occipital cortex, visual, 8-80
Occipital lobe, sensory function, 9-111
Occlusion, spinal, 9-70
Oculogyral illusion
in motion sickness, 8-59
test of vestibular function, 8-53
Oculomotor nerve (III), 9-32
Odor (see also Olfaction), 8-134 ff
Odor strength, 8-141
Odorous materials, 8-133
Oil film model membrane, 1-17

Olfaction, 8-132 ff
properties of stimuli, 8-133
Olfactometer, 8-139
Olfactory bulb, tract, 8-137
Olfactory discrimination, 8-142
Olfactory nerve, 9 32
as fiber tract, 9-1
Olfactory system
central connections, 8-137
components of, 8 134 ff
psychophysical study of, 8-139
Olfactory thresholds, 8-139
Olfactory vs limbic systems, 9 146
Oligodendroglia, 1-31
Oliguria, 5-43
Olivocerebellar tract, 9-46
Olivospinal tract (of Helweg), 9 29
Operant conditioning, 9 153
Operator genes, 1-187
Operon theory, 1-187
Oppenheim's reflex, 9 164
Opponents processes, visual, 8-114
Optic disc, 8 70
Optic nerve, 9-32
as fiber tract, 9-1
Optic nerve and tract, 8 77
Optical defects, 8-68
Optics of eye, 8-66 ff
Optokinetic nystagmus, 8-56
Orbitofrontal region, cerebral cortex, 9-117
Organ of Corti, 8-17
Organelles of cells, 1-1
diagram of relations, 1-10
structure and types, 1-2
Organic acids yield H^+, 4-103
Organification of iodide, 7-31
Orienting responses, 8-57
Orotic aciduria, hereditary, 1 156
Orthodromic vs antidromic stimulation, 1-53
Orthopnea, 6-60
Oscillation of neural circuit, 9-102
Osmium fixation for EM, 1-3
Osmogram, 8-135
Osmoreceptors, 9-126
control vasopressin release, 7 25
control water excretion, 5-18
Osmosis, 1-27
as adaptive mechanism, 1-28
Osmotic diuresis, 5-26
Osmotic fragility of RBC, 4-57
Osmotic pressure (see also Pressure, osmotic), 5-6
definition and measurement, 1-27
Osteoclasts, 7-46
Otoconia, 8-44
Otolith organs, 8-44
electrical response of, 8-59
stimulated by gravity, 8-49
Ottoson potential, 8-135
Oval window, 8-14
Ovarian compartments, 7-101
Ovarian cycle, 7-101 ff
Ovary, control of, 7-101
Overshoot of action potential, 1-46
Ovulation, 7-102
Oxalacetate
in gluconeogenesis, 7-132
in lipid synthesis, 7-124
Oxidation of amino acids, 1-132
Oxidation, water from, 4-114
Oxidative deamination, 1-132
Oxidative phosphorylation
and energy release in muscle, 1-102
and thyroid hormones, 7-39
Oximetry, 6-10
Oxygen
affinity of blood, 6-12
arteriovenous utilization during exercise, 3-122
binding with hemoglobin, 4 43
blood flow and organ consumption of, 3-123
cerebral circulation response to, 3 212
chemoreceptor effect on circulation, 3-182 ff
coefficient of utilization, 6 14
content in blood (see also Dissociation curve), 6-6
diffusion in lungs, 6-4
in measurement of cardiac output, 3-158
measurement in blood, 6 8 ff
renal consumption and blood flow, 3-237, 238
resistance to diffusion in lung, 6 24
Oxygen apnea, 6-82
Oxygen capacity of blood, 4-41
at high altitude, 6-67
in emphysema, 6-74
Oxygen capacity of hemoglobin, 6 6
Oxygen consumption
and thyroid function, 7-38
of nerve, 1-51
of salivary glands, 2-21
Oxygen debt in muscle, 1 112
Oxygen diffusing capacity, 6-24 ff
Oxygen saturation of blood, 6-6
at high altitude, 6-65
in cyanosis, 6-79
in emphysema, 6-74
in hypoxia, 6-64
in pneumonia, 6-69
in presence of shunts, 6-76
Oxygen tension
and erythropoiesis, 4-26
and multiple factor theory, 6-56
at high altitude, 6-65
Bohr integration for estimation, 6-25
effect on respiration, 6 54
in alveoli of lung, 6-22
in blood, 6 5
in hyperbaric chamber, 6 22
in hypoxia, 6-64
in lung capillaries, 6-13
in oxygen therapy, 6 82
in tissues, 6-18
influence on breathing, 6 41
Oxygen therapy, 6-81
Oxygen toxicity, 6-82
Oxygen transport in blood, 6-1 ff
Oxyhemoglobin, Bohr integration for, 6-25
Oxyhemglobin dissociation curves, 6-11 ff
abnormal hemoglobins, 6-12
effect of anemia, 6-15
effect of CO, 6-15
effect of CO_2, 6-14 ff
effect of H^+, 6-14
fetal blood, 6-12
normal and shifted positions, 4-44
of hemoglobin variants, 4-45
Oxyntic cells of stomach, 2-32
Oxytocin
chemical structure of, 7-25 ff
control of secretion, 7-27
in childbirth, 7-28
neurosecretion of, 7-24

p-Aminohippuric acid (PAH)
secretion by kidney, 5-14
tests based on clearance, 5-39
Pacemaker
gastric, 2-106
intestinal, 2-112
Pain, 8-154 ff
esophageal, 2-104
neural pathways for, 8-157
referred, 2-11
respiratory reflex from, 6-51
response to, 8-156
visceral, 2-11 ff
Pairing, purine-pyrimidine, 1-151
Pancreas, exocrine, 2-49 ff
tests of function, 2-57
Pancreatectomy, 7-76
Pancreatic cell, 1-9
Pancreatic islets, 7-76 ff
Pancreatic juice, 2-50 ff
electrolyte content, 2-50
enzymes, 2-51
Pancreatic secretion, 2-53 ff
control, 2-55
neural vs humoral control, 2-57
Pancreozymin, 2 15
control of pancreas, 2 56
evokes insulin release, 7-80
test of pancreatic function, 2-57
Paneth cells, intestinal, 2-80
Papillae of tongue, 8 123
Paradoxical cold sensation, 8-153
Paradoxical sleep, 9-142
Parahemophilia, 4 76
Parallax in depth perception, 8-118
Parallel elastic elements in muscle, 1-65
Paralysis agitans, 9-99 ff
Paralytic ileus, 2-115
Paralytic salivary secretion, 2-26
Paraplegia, 9-162
Parasympathetic control
of gastric secretion, 2-35
of pancreas, 2-56
of salivary secretion, 2-27
Parasympathetic ganglia, 9-55
Parasympathetic nervous system, 9-53 ff
summary of functions, 9-59
Parathormone (PTH)
action in intestine, 7-46
action on bone, 7 44
chemical structure, 7-43
vs calcitonin, 7-42
Parathyroid hormone (see Parathormone, PTH)
Parathyroidectomized animal, 7-44
Paraventricular nuclei, 7-25
Paravertebral ganglia, 9-55
Parietal cells of stomach, 2-33
Parkinson's disease, 9-99 ff
clinical description, 9-169
Parotid glands, 2-17
Paroxysmal hemoglobinuria, 4-59
Paroxysmal nocturnal dyspnea, 6-60
Partial pressures of gases, 6-3
Patellar reflex, 9-72
Patterning, neural, 9-12
Patterns of sensory impulses, 8-7
Pauling-Corey model of base pairs, 1-152
Pauling oxygen analyzer, 6-8
Pavlov, Ivan, 9-151
Peduncles, cerebellar, 9-46 ff
Pelvic nerve, 9-58
Pelvis, renal, 5-52
Penetrance of genes, 1-137
Penicillin, renal secretion, 5-13
Pentose phosphate pathway, 7-126
Pepsin in gastric juice, 2-37
Pepsinogen cells of stomach, 2-31
Peptidase, pancreatic, 2-52
Peptide bond formation, 1-180
Peptide bonds in protein, 1-121
Peptide chain initiation, 1-179
control of, 7-153

Peptide chain termination, 1-181
Peptide chains in ribosomes, 7-152
Peptides
absorption of, 2-90
in protein synthesis, 7-148
Perception of depth, 8 117
Perception of rotation, 8-53
Perceptual duration, visual, 8-112
Pericardium, anatomy, 3-27
Perilymph
and sound transmission, 8-14
electrolytes of, 8-18
Periodic breathing, 6-58
Periodicity in muscle fibers, 1-64
Peristalsis
esophageal, 2-100
gastric, 2 106
Peristalsis in small intestine, 2 113
Permeability coefficient, 1-18
Permeability of membrane
change during nerve impulse, 1-49
of muscle fiber, 1-93
of nucleus, 1-7
Permease, 1-139
in sugar transport, 1-5
Pernicious anemia, 4-33
and gastric secretion, 2-38
pathogenesis, 4-37
Peroxidase in thyroid gland, 7-32
Peroxisomes and ER, 1-11
Perspective, visual, 8-117
Petit mal epilepsy, 9-171
EEG pattern, 9-108
Phage, RNA sequences, 1-178
Phagocytosis by granulocytes, 4-62
Phantom limb, 9-167
Pharyngeal peristalsis, 2-101
Phases of gastric secretion, 2-41 ff
Phases of ovarian cycle, 7 104
Phenotype vs genotype, 1-137
Phenylketonuria (PKU), 1-141
Pheochromocytoma, 7-75
Philadelphia chromosome in leukemia, 4-19
Phillipson's reflex, 9-73
in decerebrate posture, 9 92
Phloridzin, effect on blood glucose, 7-83
Phlorizin, and glucose transport in kidney (see also Phloridzin), 5-10
Phosphate
and acidification of urine, 5-30
binding to hemoglobin, 4-44
in renal dysfunction, 5-47
PTH and exchange in bone, 7-45
PTH and renal loss of, 7-45
source of H^+, 4-103
transport in kidney, 5-13
Phosphate bond energy
and fat absorption, 2-91
and HCl secretion, 2-40
and muscle contraction, 1-96
Phosphate buffer system, 4-105
Phosphate transport in plasma, 7-43
Phosphaturia, 7-45
Phosphocreatine in muscle, 1-96
Phosphodiesterase
in second messenger theory, 7-5
relation to lipolysis, 7-129
Phosphofructokinase, 7-120
Phosphogluconate pathway, 4-28
Phospholipids in membranes, 1-16
Phosphorus of plasma, 4-2
Phosphorylase, 7-118
Phosphorylases of muscle, 1-99
Phosphorylation
and energy release in muscle, 1-102
and thyroid hormones, 7-39
in activation of lipase 7-130
in control of lipogenesis, 7-141
Phosphorylation of ADP, 1-95
Phosphorylation of glucose
blocked by growth hormone, 7-145
in muscle, 7-116
Photochemistry of vision, 8-81
Photoelectric oximetry, 6-10
Photometric units, 8-65
Photons, 8-64
Photopic curve of luminosity, 8-65
Photopigments, 8-84
Physics of sound, 8-10
Physiological capacity of bladder, 5-53
Pigment of eye, 8-70 ff
Pigmentation in Wilson's disease, 9-169
Pigmentation of skin and cyanosis, 6 81
Pigments of bile, 2-64 ff
Pigments of cones, 8-82
Pinocytosis, 1-25
in blood capillaries, 4-83
in intestinal absorption, 2-90
Pitressin
decrease in coronary blood flow, 3-139
renal blood flow effects of, 3 236
Pituitary-adrenal axis, 7-61 ff
Pituitary depletion assay, 7-16
Pituitary gland (see also Adenohypophysis, Neurohypophysis)
embryology of, 7 24
Frohlich's syndrome, 2-2
hypothalamic connections, 9-41
Pituitary-thyroid axis, 7-36
Placenta
blood flow in, 3-239
development of, 7-107
estrogen synthesis in, 7 110
lactogen secretion, 7-109
secretion of steroid hormones, 7-103
thyrotropin secretion, 7 109
Placing reactions, 9-113
Planes of semicircular canals, 8-41
Plantar reflexes, 9-164
Plasma
alkali reserve, 6-9
calcium concentrations, 7-43
composition of, 4-2
phosphate concentrations, 7-43
Plasma amino acids
affected by insulin, 7-84
effect on growth hormone release, 7-143
Plasma cells, 4-94
Plasma constituents (see also name of constituent)
and renal function, 5-8
in acid-base disturbances, 4 110
Plasma cortisol, 7-53
assay of, 7-53
Plasma flow, renal, 5-15 ff
Plasma iodine, 7-33
Plasma proteins, 4-3
and edema formation, 4-88
as buffer system, 4-105
binding to testosterone, 7-97
biosynthesis and dynamic state, 4-7
osmotic pressure, species comparison, 3-4
variations in, 4-6
Plasma steroid hormones, 7-103
Plasminogen vs fibrin, 4-77
Platelet formation, 4-18
Platelet transfusion, 4-74
Platelet turnover rate, 4-74
Platelets (see Blood platelet)
Pleated sheet structure of protein, 1-126
Pneumonia
lobar vs broncho-, 6-69
oxygen therapy in, 6-81
Pneumotaxic center, 6 45
location, 6-42
Pneumothorax
effect on interpleural pressure, 6-38
effect on mechanoreceptors, 6 48
Poikilothermic animals, 9 128
Poiseuille's law, 3-23, 24
Poisoning
carbon monoxide, 6 78
cyanide, 6-78
Polar conduction in intestine, 2 115
Polar properties of protein, 1-126
Polarization pattern in saccule, 8-48
Polarography for O_2 tension, 6-10
Polycythemia, 6-80
Polymerase
for RNA synthesis, 1-171
in synthesis of nucleotides, 1-150
Polymerization of actins, 1-65
Polymers in cells, 1-1
Polypeptides
assembly into proteins, 1-183
elongation of, 1-179
Polyribosomes and mRNA, 1-182
Polyuria, 5-42
Pool, motor neuron, 9-71
Pools of amino acids, 1 131
Pools of blood granulocytes, 4-67
Pores in membrane
and transport rate, 1 4
of erythrocyte, 1-17
Porphyrias, 4-47
Porphyrin enzymes, 4 39
Porphyrin of hemoglobin, 4-41
Porphyrin synthesis, 4-41
Porphyrins from hemoglobin degradation, 4-35
Portal system, hypothalamico-hypophysial, 7-13
Position in space and vestibular function, 8-40
Positive control by operon, 1 189
Positive pressure breathing, 6-49
Postcentral gyrus, 9-109
Posterior lobe of cerebellum, 9-43
Posterior lobe of pituitary gland (see also Neurophyophysis), 7-24 ff
Posterior roots, spinal, 9 23
Postganglionic fibers, 9-55 ff
peripheral distribution, 9-56
Postsynaptic inhibition, 9-79
Post-translational controls of genes, 1-192
Postural reflexes, 9-95 ff
Posture
and stretch reflex, 9-72
as rhythmic mechanism, 9-102
control of, 9 88 ff
Potassium current
during action potential, 1-49
transmembrane in muscle, 1-92
Potassium depletion in alkalosis, 4-110
Potassium equilibrium potential, 1-92
Potassium ions
active transport processes, 1-23
affect secretion of H^+, 5-34
aldosterone and tubular transport, 7-59
and cochlear potentials, 8-22
and Hodgkin-Huxley theory, 1-36
and hormone secretion, 7-21
and insulin release, 7-80
and transmembrane potential, 1-34
depletion, with polyuria, 5-42
in vestibular function, 8-40
renal transport, 5-26, 27

Potential (see also name of potential)
Potentials, electrical
and HCl secretion, 2-40
and neurosecretion, 7 14
and salivary secretion, 2-27
cochlear microphonic, 8-20
during conditioning, 9-151
during nerve impulse, 1-48 ff
excitatory postsynaptic, 9 76 ff
from motor unit in muscle, 1 105
in association cortex, 9-110
in olfactory system, 8 135
in spinal roots, 9-11
in vestibular system, 8-57
in visual system, 8-86 ff
inhibitory, 9-77 ff
of cerebral cortex, 9-105
of end plates, 9-13
of small intestine, 2-112
of stomach, 2 106
principles of bioelectricity, 1 32 ff
related to hearing, 8-22
to sensory stimuli, 8-6
Potentials, electrochemical, 1 21
in active transport, 1 21
in carrier transport, 1-21
Potentials, electrotonic
as CES, CIS, 9-14
during stimulation of nerve, 1-40
of auditory hair cells, 8-19
volume conduction, 9 11
Potentials, resting vs action, 1-34
Potter turbinometer, measurement of blood flow, 3-155
Precentral motor cortex, 9-89
Prefrontal region, cerebral cortex, 9 117
Preganglionic fibers, 9-55 ff
Pregnancy
and Rh antigen system, 4-51
and steroid hormones, 7-103
effect on blood volume, 4-120
hormonal control, 7-107 ff
vomiting in, 2-110
Pregnenolone, 7-51
Preservation injury in blood, 4-53
Pressure, hydrostatic
and lymph formation, 4-85
and Starling principle in capillaries, 4-122
during swallowing, 2-99
gastroduodenal, 2-107
in ileum, 2-116
in urinary tract, 5-53
measurement in capillaries, 4-86
relations in kidney, 5-6
Pressure, osmotic, 1-27
and glomerular filtration, 5-6
and lymph formation, 4-85
and Starling principle in capillaries, 4-122
variations affect lymph flow, 4-87
Pressure-volume relation
for gas bubble, 6-39
for lung, 6-37
Pressures that move air, 6-34 ff
Pressures that move fluid, 4-123
Presynaptic inhibition, 1-56
in spinal reflexes, 9-80
Primaquine causes G6PD deficiency, 4-29
Primates, cortical lesions, 9-114
Priority in control systems, 9-17
Prism for light, 8-65
Problem-solving tests, 9-155
Progesterone
in ovarian cycle, 7-105
in synthesis of glucocorticoids, 7-51
metabolic pathways of, 7-104
secreted during pregnancy, 7-107
Progestin
in reproductive cycle, 7-103
mechanism of action, 7-107
transport and metabolism, 7-106
Proinsulin
model of protein assembly, 1 184
synthesis of, 7-77
Projectile vomiting, 2-110
Prolactin inhibitory factor (PIF), 7 15
uncertain chemical nature, 7 19
Proliferation, calculation of, 4-15
Promoter in operon, 1-189
Proprioceptors, 8-1
Prostaglandins and body temperature, 9-136
Prostate gland, androgen on, 7 99
Protein biosynthesis
affected by steroid hormones, 7-7
and memory trace, 9-159
and thyroid hormone action, 7-39
codons for, 1-175
function of RNA, 1-179
hormonal control of, 7 147
inhibition by antibiotics, 1-181
Protein biosynthesis in liver, 7-151
Protein-bound iodine (PBI), 7-34
Protein metabolism, 1-129
and adrenal hormone actions, 7-56
and gluconeogenesis, 7-136
hormonal control of, 7 147
Protein molecules
conformation of, 1-124
models for their assembly, 1-184
types of bonds in, 1-122
Protein sparing compounds, 1-130
Proteins
absorption of, 2-89
assembly from polypeptides, 1-183
classes of, 1-115
physicochemical properties, 1-121
summary of functions, 1-114
Proteins in bile, 2-69
Proteins in interstitial fluid, 4-85
Proteins in lymph, 4-85
Proteins in plasma (see also Plasma proteins), 4-3
Proteinuria
absent in pyelonephritis, 5-41
in glomerulonephritis, 5-40
in nephrosclerosis, 5-41
in nephrotic syndrome, 5 40
significance of, 5-41
Proteoses, absorption of, 2-90
Prothrombin, 4-75
Prothrombin time, 4-78
Protons (see also Hydrogen ions), 2-39
Protoporphyrin IX, 4-41
PRPP-amidotransferase, 1-147
PRPP in gout, 1-155
PSP (phenol red)
renal secretion, 5-13
test for excretion, 5-39
Psychological sex differentiation, 7-94
Psychomotor epilepsy, 9-171
Psychophysical study (see Techniques, psychophysical),
Pulmonary artery, "edge" pressure of, 3-206
Pulmonary blood vessels, reflex from distension of, 6-51
Pulmonary edema, dyspnea in, 6-60
Pulmonary veins, critical closing pressure of, 3-207
Pulmonary ventilation
abnormal patterns, 6-58
as rhythmic mechanism, 9-103
control of, as regulation, 9-125
efficiency, 6-37
first breath of infant, 6-40
hyperpnea during exercise, 6-57
hysteresis in curves, 6-38
interpleural pressures, 6-38
multiple factor theory of control, 6-56
pressure-volume relation, 6-37
pressures for air flow, 6-34 ff
range of normal values, 6-33
regulation of, 6-52 ff
regulation of $[H^+]$, 4-108
resistance to air flow, 6-35
significance of dead space, 6-20
total vs alveolar, 6-20
volume-flow relations, 6-34
work of breathing, 6-36
Pulmotor, 6-84
Pulvinar, 9-40
Pumps, ionic, in nerve, 1 49
Pupil
optics of, 8 67
properties of, 8-71
Purine derivatives, 1-142
Purine-pyrimidine pairing, 1-151
Purines
biosynthesis of, 1 147
deamination of, 1 134
degradation of, 1 151
in genetic code, 1 175
incorporation into nucleic acids, 1-150
reutilization of, 1-149
synthesis in gout, 1-154
Purkinje cells
conductance of impulse, 3-57
of cerebellum, 9-44
Purkinje fibers, transmembrane potential, 3-49
Puromycin
inhibits memory 9-159
inhibits protein synthesis, 1-182
Purpura, 4-80
Putamen, 9-50
Pyelonephritis, 5-41
Pyloric glands, 2-32
Pylorospasm, 2-109
Pylorus of stomach, 2-104
Pyramid of cerebellum 9-99
Pyrimidine derivatives, 1-142
Pyrimidines
biosynthesis of, 1 143
deamination of, 1-134
degradation of, 1-151
in genetic code, 1-175
incorporation into nucleic acids, 1 150
reutilization of, 1 149
Pyrogen
action on kidney, 5-16
on hypothalamus, 9-136
Pyruvate dehydrogenase
and control of gluconeogenesis, 7-138
in glycolytic unit, 7-120
Pyruvate oxidation, 1-103

Quadriplegia, 9-162
Quanta of radiation, 8-64
Quantal nature of transmission, 1-58
Quantum theory of light, 8 63
Quaternary structure of protein, 1-128

Radiation, heat loss, 9-129
Radioactive isotopes
Fe, Cr, for RBC volume, 4 119
for transmembrane flux, 1-25
^{131}I for plasma volume, 4-118
Radioautography
of anterior pituitary cells, 7-12
with tritiated thymidine, 4-13
Radioimmunoassay, 7-3
for ACTH, 7-62

Radioimmunoassay—*continued*
of calcitonin, 7-48
of corticosteroids, 7-54
of insulin, 7-76
Rage
hypothalamic, 9-145
sham, 9-144
Rahn sampler for gases, 6-4
Rami communicantes, 9-56
Ramon y Cajal, 9-44
neuron doctrine, 9-68
Rapid eye movement (REM), 9-138
Rathke's pouch, 7-24
Ratio, HCO_3/CO_2, 4-105
Reactions, placing and hopping, 9-113
Reactive hyperemia, 3-221
Reading and speech, 9-148
Reading frame shifts, 1-185
Receptive fields of retina, 8-89
Receptive fields of skin, 8-147
Receptor potentials of retina, 8-86
Receptor sites
for catecholamines, 7-73
for gastrointestinal hormones, 2-16
for TSH, 7-36
in autonomic system, 9-63
in nucleus for testosterone, 7-99
Receptors, postsynaptic, 1-52
Receptors, sensory, 8-1
for olfaction, 8-134
for pain, 8-154
for taste, 8-123
for temperature, 8-152
for touch, 8-145
in retina, 8-74
in vestibular system, 8-44
Recessive genes, 1-136
Reciprocal innervation, 9-73
Recognition of codons, 1-177
Recording, electrical, 9-9 ff
Recovery from neural deficits, 9-6
Recovery heat in muscle, 1-87
Recruitment
auditory, 8-35
motor units, 1-81
Rectal motility, 2-119
Rectal temperature, 9-128
Rectification in muscle, 1-92
Rectum
defecation, 2-119
reflex control, 9-163
Red blood cells (see Erythrocyte)
Red muscle fibers, 1-109
Red nucleus, 9-51
in decerebrate state, 9-93
Red vs white muscle, 9-90
biochemistry of, 7-112
Referred pain, 8-155
from gut, 2 11
Reflex (see also name of reflex)
swallowing, 2-96
Reflex arc, 9-67
Reflex secretion of saliva, 2-25
Reflexes, classification, 9-68
Reflexes, feeding, 2-3
Reflexes, respiratory, 6-48 ff
cause of vasoconstriction in finger, 3-187
from baroreceptors, 6-48
from CO_2 detectors, 6-52
from H^+ detectors, 6-56
from low O_2 detectors, 6-54
from pulmonary vessels, 6-51
from thermoreceptors, 6-52
Hering-Breuer, 6-47
miscellaneous types, 6-51
pain, 6-51
Reflexes, spinal
during recovery from shock, 9-163
electrophysiology of, 9 74 ff
Reflexes, urinary, 5-51
Reflexes, visceral, 2-9 ff
Refraction of light, 8-64
Refractive index of eye, 8-66
Refractory periods
of muscle, 1-80
of nerve, 1-45
Refractory states and ion pumps, 1-50
Regeneration of nerve fibers, 1-31
Regional potentials, cochlear, 8-22
Regulation, 9-18
genetic, 1-186 ff
metabolic, 7-112
of acid-base balance, renal, 5-27 ff
of blood volume, 4-117
of body fluids, 4-115
of body temperature, endocrine, 9-136
of body temperature, neural, 9-128 ff
of erythropoiesis, 4-25
of granulocytopoiesis, 4-69
of $[H^+]$, 4-106
of platelets, 4-74
of sodium excretion, 5 24 ff
of ventilation, 6-52 ff
of ventilation during exercise, 6-57
schemata, 9-122
typical, 9-121
Regulation vs control, 9-18
Reinforcement in conditioning, 9-154
Reissner's membrane, 8-17
Rejection of grafts, 4-98
Relative refractory period
of muscle, 1-80
of nerve, 1-45
Relaxing factor, 1-93
Releasing factors 7 13
activity in human, 7-20
assay of, 7-15
chemistry of, 7-18
clinical significance of, 7-23
in blood of portal system, 7-13
mechanism of action, 7 20
REM (rapid eye movement), 9-138
Renal (see also Kidney, Glomeruli, Tubules, name of process, name of substance excreted)
Renal circulation, 3-227 ff
Renal erythropoietic factor, 4-26
Renal failure, 5-43
effect on acid-base balance, 4-106
Renal glycosuria, 5-46
Renal hypophosphatemia, 5-46
Renal tubular acidosis, 5-45
Renin
bioassay for, 7-62
control of Na^+ secretion, 5-25
Renin-angiotensin, vasoconstrictor effects of, 3-172
Renin-angiotensin system, 7-62
Rennin in gastric juice, 2-37
Renshaw cell inhibition, 1-57
in spinal reflexes, 9-81
Replicases for RNA, 1-171
Replication of DNA, 1-163
Replication of RNA, 1-169
Repression of enzymes, 1-187
Repressor vs inducer, 1-187
Reproductive cycle, female, 7-101
Residual volume of lung, 6-31
Resistance
to diffusion of CO_2 in lung, 6-27
to diffusion of O_2 in lung, 6-24
to pulmonary air flow, 6-35
Resonant frequency and odor, 8-143
Respiration (see also Inspiration, Expiration, Pulmonary ventilation)
Respiration, artificial, 6-84
Respiration, control of, 6-41 ff
multiple factor theory, 6-56
Respiration, muscles of, 6 28 ff
abdominals, schema of control, 6 51
maximal pressures developed, 6-33
Respiration, symbols used, 6 1 ff
Respirators, 6 84
Respiratory centers
automaticity of, 6-44
identification of, 6-42 ff
recording activity from, 6-45
Respiratory distress syndrome, 6-40
Respiratory gases (see also CO_2, O_2), 6-1 ff
Respiratory minute volume, 6 52 ff
Respiratory quotient
of nerve, 1-51
of salivary glands, 2-21
Respiratory rate vs dead space, 6-21
Respiratory reflexes (see Reflexes, respiratory)
Respiratory sinus arrhythmia, 3-187
Rest-length of muscle, 1-63
Restiform bodies, 9-46
Resting membrane potential, 1-89
Resting tension of muscle, 1-82
Resting vs action potentials, 1-34
Restraining mechanism, hypothalamic, 9-144
Resuscitation, 6-83 ff
Reticular formation, medullary, 9-31 ff
activating mechanism, 9-140
influence on visual system, 8-81
respiration, control of, 6-42 ff
Reticular formation, spinal, 9-22
Reticulocytes, 4-31
in erythrocytopoiesis, 4-24
Reticulospinal tract, 9-29
Reticulum
endoplasmic, 1-7
sarcoplasmic (see Sarcoplasmic reticulum)
Retina, 8-72
absolute sensitivity of, 8-84
distribution of rods, cones, 8-69
interaction of units, 8-87
organization of, 8-75
receptive fields, 8-89
receptive fields for color, 8-115
Retinal disparity, 8-120
Retinaldehyde, 8-82
Reverberatory network, 9-76
Reversion of mutation, 1-185
Reynolds' number, 3-7, 19
for air flow in lung, 6-35
Rh antigen system, 4-51
Rheological properties of blood, and suspension stability, 4-61
Rhinencephalon
and limbic system, 9-146
and olfaction, 8-138
Rhodopsin, 8-81
Rhythmic mechanisms, neural, 9-102
Rhythmic segmentation, 2-111
Rhythmicity of bowel, 2-113
Ribonuclease
denaturation of, 1-128
pancreatic, 2-52
Ribonucleic acid (see RNA)
Ribonucleotide biosynthesis, 1-146
Ribose in RNA, 1-142
Ribosomal RNA
and nucleolar function, 1-6
structure and function, 1-172
Ribosome cycle, 7-152
Ribosomes
aggregation of, 7-152

hormonal control of protein synthesis, 7-148
in hypophysectomy, 7-149
significance of RNA, 1 172
Ribs, in respiration, 6 29
Rickets, vitamin D resistant, 5-46
Rifamycin vs RNA synthesis, 1-171
Righting, natural, 9-96
Righting reflexes, 9 96
of decorticate animal, 9-94
Rigidity
decerebrate, 9-92
parkinsonian, 9-100
Rigor of muscle, 1-75
RNA (see also Nucleic acids)
messenger (see Messenger RNA)
role in cardiac hypertrophy, 3 45 ff
RNA and learning, 9-159
RNA biosynthesis, 1-169
and corticosteroid actions, 7-57
and thyroid hormone action, 7-39
in bone, 7 45
stimulated by androgen, 7-99
RNA molecules, 1-169
structure of single strand, 1-192
RNA of nucleolus, 1-6
RNA polymerase, 1-169
control via, 1-190
RNA replicases, 1-171
RNA sequences of phage, 1-178
RNA structure, 1-142
RNA vs DNA, 1-142
Rods vs cones
distribution of, 8-69
in dark adaptation, 8-101
microscopic anatomy, 8-74
Rollbewegungen, 2-114
Rossolimo's reflex, 9-87
Rotameter, measurement of blood flow, 3-155
Rotating environment, 8-54
Rotation, perception of, 8-53
Rough ER, 1-8
Round window, 8-14
Rubrospinal tract, 9-29
connections of, 9-50
Ruminants, urea cycle, 5-34

Saccadic eye movements, 8-97
Saccules, 8-46
Sacral autonomic outflow, 9-58
Saline infusions, 4-117
Saliva
composition, 2-22
functions, 2-28
Saliva and thirst, 2-6
Saliva, secretion of, 2-21
disturbances, 2-30
paralytic, 2-26
rate and volume, 2-22
Saliva spreading and heat loss, 9-130
Salivary enzymes, 2-29
Salivary glands, circulation at rest vs maximum vasodilatation, 3-166
Salivatory nuclei, fibers from, 9-57
Salivatory nucleus (inferior), 9-37
Salivatory nucleus (superior), 9-36
Salt (see Electrolyte, name of salt)
Salty taste, 8-122
Sarcolemma
relation to T tubule, 1-71
stimulation of muscle via, 1-90
Sarcomere
definition, 1-62
myocardial, 3-28 ff, 88 ff
Sarcoplasm of muscle, 1-62
Sarcoplasmic reticulum, 1-70, 3-31
as relaxing factor, 1-94
calcium storage and release, 3-36
concentration of calcium by, 1 90
stores calcium ions, 1 94
Satiety, defined, 2-2
Saturation of carrier transport, 1-20
Scarpa's ganglion, 8-49
Schafer's method, resuscitation, 6-83
Scholander method for gas analysis, 6 8
Schwann cells, 1-31
Scleroderma and dysphagia, 2-103
Sclerosis, disseminated, 9-170
Scotopic curve of luminosity, 8-65
Seasickness, 2-110
Second messenger theory, 7-5
Secondary sex characters, 7-94
Secretin, 2-15
and secretion of bile, 2-70
control of pancreas, 2 56
evokes insulin release, 7 80
test of pancreatic function, 2-57
Secretion
by colon, 2-82
by pancreas, 2-53 ff
by renal tubules, 5-13 ff
by small intestine, 2-79
gastric, 2-31 ff
gastric phases, 2-44
of bile, 2-60 ff
of glucagon, 7-79
of H^+ in kidney, 5-29
of HCl, 2-38
of insulin, 7-79
of saliva, 2-21
of saliva, disturbances of, 2-30
Secretion granules, anterior pituitary gland, 7-9
Secretory capacity of stomach, 2-45
Secretory vesicles, adrenal medulla, 7-72
Sedimentation of blood, 4-60
Segmental nerves, 9-23
Segmentation of bowel, 2-111
Seizures, epileptic, 9-172
Selectivity for frequency, 8-28
Self-stimulation of brain, 9-154
Semicircular canals, 8-41
Semilunar valves, closure, mechanism of, 3-107
Seminiferous tubules, 7-99
Semipermeable membranes, 1-28
Sensation
consciousness of, 8-2
in control of food intake, 2-4
visceral, and digestion, 2-10 ff
Sensitivity of receptors
auditory, 8-19
olfactory, 8-140
retinal, 8-84
thermal, 8-152
Sensorimotor cortex, cerebral, 9-110
Sensory changes in CNS lesions, 9-162
Sensory coding, 8-3
Sensory disabilities, 9-24
Sensory nerves, 9-74
Sensory pathways, 9-24 ff
Sensory processes, 8-1 ff
Sensory responses, neural, 8-6
Septal defects, cardiac, cause hypoxia, 6-77
Series elastic element, 1-81
Serotonin
and inflammation, 4-64
as transmitter, 1-59
distribution in CNS, 9-83
in small intestine, 2-80
Serous cells, salivary, 2-19
Sertoli cells, 7-100
Serum proteins, 4-4
Setpoint for body temperature, 9-136
Setpoint of control system, 9 19
Sex characters, secondary, 7-94
Sex chromosomes, 1 137
abnormal numbers of, 7-88
and sex differentiation, 7 88
Sex differentiation, 7-88 ff
stimulated by androgen, 7-99
Sham rage, 9 144
Shock (see also Spinal shock, etc)
cardiac blood flow during, 3-141
electric, and memory, 9 158
Shortening of muscle, 1 80
Shortening reaction, 9-92
Shunts, pulmonary, 6-22 ff
causes of hypoxia, 6-76
Sialorrhea, 2-30
Sickle cell anemia
abnormal hemoglobin, 4-31
as a proteinopathy, 1-138
hemoglobin variants, 4-37
Siggaard-Andersen nomogram, 4-112
Sign of Babinski, 9-161
diagnostic value, 9-164
Signal generation, neuronal, 9-16
Signals, neural, coding of, 9 12
Signet ring cells of anterior pituitary gland, 7-9
Signs vs symptoms, neural, 9-4
Silent period of reflex, 9-86
Silicone prevents clotting, 4-76
Silver chromate technique, 1-29
Single unit analysis, 9-10
of cochlear function, 8-24
of lateral geniculate body, 8-92
of olfaction, 8-135
of retinal interactions, 8-87
of taste, 8-127
Sinoatrial node, 3-55
Sinus node, heart, conductance velocity, 3-55
Skeletal muscle
arteriovenous anastomoses in, 3-222
blood circulation in, 3-217
blood pH effects on circulation, 3-218
bradykinin vasodilation effect in, 3-219
carbon dioxide effects on blood flow in, 3-218
circulation at rest vs maximum vasodilatation, 3-166
collateral blood flow, 3-222
contraction, compared with myocardium, 3-36
control of, 9-66 ff
exercise effect on blood flow in, 3-220
neural control of blood flow in, 3-217
reactive hyperemia, 3-221
red vs white, 9-90
tone of, 9-88
tone vs spasticity, 9-166
Skin
blood circulation in, 3-213 ff
central neural control of blood flow, 3-216
circulation at rest vs maximum dilatation, 3-166
dermographism, 3-215
dorsal root vasodilator fibers, 3-216
flare reaction, 3-214
flush reaction, 3-215
local edema, 3-215
red reaction, 3-215
venous circulation of, 3-214
wheal reaction, 3-215
white reaction, 3-215
Skin and touch sensation, 8-147

Skin pigment and cyanosis, 6 81
Skin temperature, 9-135
Sleep, 9 137 ff
  humoral control, 9-142
  paradoxical, 9 142
  significance of, 9-143
Sleeping sickness, 9-141
Sliding filament theory, 1 72
Slow muscle fibers, 1-110
Slow waves, EEG, 9-138
Small intestine
  nerve supply, 2-9
  secretion by, 2 79
Smell, 8-132 ff
Smooth ER, 1 8
Smooth muscle, control of, 9-53 ff
Smooth pursuit eye movement, 8 98
Snake venom and hemolysis, 4-60
Snellen chart, 8 107
Sodium chloride and thirst, 2-6
Sodium current
  during action potential, 1-48
  transmembrane in muscle fiber, 1-93
Sodium cyanide and respiration, 6-56
Sodium inactivation, 1-48
Sodium ions
  active transport processes, 1 23
  aldosterone and tubular transport, 7-59
  and glucose absorption, 2-87
  and Hodgkin-Huxley theory, 1-36
  and Meniere's disease, 8-61
  and size of action potential, 1-46
  in countercurrent system, 5-20
  renal transport, 5-23 ff
Sodium pump in nerve, 1 49
Soleus muscle, red, 1-109
Solubility coefficients of gases, 6-3
Somatotrophin (see Growth hormone, GH)
Somesthetic area, cerebral, 9-109
Somnolence, 9-141
  with hypothalamic lesion, 9-167
Sound
  physics of, 8 10
  stereophonic, 8-33
Sound level and deafness, 8-34
Sound localization, 8-33
Space station, 6-69
Spastic hemiplegia, 9-116
Spatial factors, visual, 8-105
Spatial modulation transfer function, 8-108
Spatial resolution in eye, 8-69
  and clarity of vision, 8-105
Spatial summation, 9-70
Specific dynamic action, 1-132
Specific gravity of urine, 5-21
Specific heat of blood, 4-1
Specific taste fibers, 8-126
Spectral luminosity function, 8-65
Spectral response function, 8-82
Spectrum of electromagnetic radiation, 8-64
Speech, 9-147 ff
Speech and hearing, 8-29
Speech audiometry, 8-38
Speech centers, 9-120
Speech disorders, 9-148
Spermatogenesis, 7-99
Spherical aberration, 8-67
Spherocytosis, hereditary, 4-36
Sphincter, esophageal
  lower, 2-102
  upper, 2-98
Sphincter of oddi, 2-75
Sphincters, urinary, 5-49
Spinal accessory nerve, 9-38
Spinal animal, 9-69
Spinal cord, 9 21 ff
  control of vasoconstriction, 3-168
  myelination of fiber tracts, 1-31
  neural tracts, 9-24 ff
  typical lesions, 9-162
Spinal nerve roots, 9-23
Spinal shock, 9-7
  effect on posture, 9-90
  stages of recovery, 9-162
Spinal state and posture, 9-90
Spinothalamic pathways, 9-25
Splanchnic circulation, 3 224 ff
Splanchnic nerves, 2-8
Splay of glucose titration curve, 5 12
Spleen
  and circulation of lymphocytes, 4-97
  blood flow, 3-225
  destruction of erythrocytes, 4-34
  forms bile pigments, 2 65
  in marrow transplantation, 4-18
Split-brain technique, 8-5
Spontaneous movement, 9-16
-S-S- bridges in insulin, 7 76
St Vitus's dance, 9 169
Standard reference levels for sound, 8-12
Stapes, functions, 8 14
Starling curves, cardiac output, 3-125
Starling principle (capillary), 4-122
  applied to glomeruli, 5-4
Starvation and gluconeogenesis, 7-134
Static reactions, neural, 9-95
Static vs dynamic receptors, 8-44
Statoconia, 8-44
Stem cell pool, 4-15
Stem cells
  in erythropoiesis, 4-21
  in granulocytopoiesis, 4-65
  in hemopoiesis, 4-12
Stepping, spinal, 9-103
Stereocilia
  auditory, 8-19
  vestibular, 8-42
Stereophonic sound, 8-33
Stereotaxic surgery, 9-5 ff
Steroid hormones
  enzyme induction by, 1-193
  of adrenal cortex, 7 50 ff
  of ovary, 7-103
  of testis, 7-95
  penetration of membranes, 7-5
Steroids (see Adrenal steroids, Ovary, Testis)
Sticky point in RNA, 1-180
Stiles-Crawford effect, 8-68
Stimulation (see Electrical stimulation, Thermal stimulation, etc.)
Stimuli, sensory, 8-1
Stimulus for olfaction, 8-132
Stimulus for pain, 8-154
Stimulus for touch, 8-145
Stomach (see also under Gastric)
  control of emptying, 2-108
  electrical activity of, 2-106
  glands of, 2-31
  motility of, 2-104 ff
  nerve supply, 2-9
Stop signals in RNA synthesis, 1-171
Storage compartment in hemopoiesis, 4-17
Storage of blood, 4-53
Strength-duration curve, 1-38
Streptomycin
  disturbs vestibular function, 8-61
  inhibits protein synthesis, 1-182
  resistance to, 1-182
Stretch reflex, 9-72 ff
  and muscle spindles, 9 82
  clinical use, 9-87
  in decerebrate state, 9-93
Striate cortex, 9-111
  visual functions of, 8 80
Striated muscle (see Muscle, skeletal)
Striations of muscle, 1-62
Stricture, esophageal, 2 103
Structural genes, 1-187
Strychnine and inhibition, 1-56
Subliminal fringe, 9-71
Subliminal stimulus for nerve, 1 37
Sublingual glands, 2-18
Submaxillary glands, 2-17
Submucous plexus, 2-8
Substantia nigra, 9-50
  dopa and melanine, 9-101
  effect of lesions, 9-100
Substrate availability, 7-134
Substrate oxidation in man, 7-125
Substrates for gluconeogenesis, 7-135
Subthalamic body (of Luys), 9-50
Succus entericus, 2-80
Suckling and oxytocin release, 7-27
Sucre piqûre, 9-65
Sucrose absorption, 2-88
Sudden death with asthma, 6-73
Sugars, transport into muscle, 7 115
Sulfate ion
  from amino acid metabolism, 1-131
  transport in kidney, 5 13
Sulfonamide drugs vs carbonic anhydrase, 6-17
Sulfonphthalein, renal secretion, 5-13
Sulfur containing amino acids, 1 133
Sulfur oxidation, 4-103
Summation
  in nerve, 1-39
  spatial, 9-70
  temporal, 9-76
Summation potential, cochlear, 8-22
Supplementary motor area, 9 113
Suppression of mutation, 1-185
Suppressor mechanism, hypothalamic, 9-141
Supraoptic nuclei, 7-25
Supraopticohypophysial tract, 7-28
Surface activity in lung, 6-38
Surgical procedures
  for deafness, 8-34
  in study of sensation, 8 5
  stereotaxic operations, 9-5
Surveillance, immune, 4-98
Sustentacular cells of testis, 7-100
Swallowing, 2-28
  deficits of, 2-97
  mechanisms of, 2-95
  stages of, 2-96
Sweating
  acetylcholine stimulation of, 9-62
  and cutaneous heat loss, 9-130
  and water balance, 4-115
  control of, 9-131
Sweet taste, 8-122
Sydenham's chorea, 9-169
Symbols for respiratory physiology, 6-1 ff
Sympathetic control
  of gastric secretion, 2-35
  of pancreas, 2-55
  of salivary secretion, 2-27
Sympathetic ganglia, 9-55
Sympathetic nervous system, 9-53 ff
  summary of functions, 9-59
Sympathin, 9-62
Symptoms vs. signs, neural, 9-4
Synapses, 1-59 ff
  general properties, 9-1
  in retina, 8-77
Synaptic delay, central, 9-75

Synaptic transmission, 1-52 ff
Syndrome
chorda tympani, 2-30
Dejerine and Roussy, 9-167
Dubin-Johnson, 2-67
Kluver Bucy, 9-145
Synthetase, 1-174
Syphilis, tabes in, 9-165
Syringomyelia, 9-165
System of cell renewal, 4-10

T-1824 for plasma volume, 4 118
Tabes dorsalis, 9-165
Tachycardia, 3 137
Taste, 8-121 ff
theories of, 8-131
Taste buds, 8-123
Taste discrimination, 8-129
Taste fibers, 8-126
Taste receptors, 8-130
Taste threshold, 8-128
Taurine in bile salts, 2-62
Techniques, biochemical
differential centrifugation, 1-172
for transmembrane flux, 1 25
Techniques, blood
antiglobulin (Coombs) test, 4 50
electrophoresis of plasma protein, 4 4
for blood volume, 4-119
for circulating RBC volume, 4-119
for clotting, 4-78
for erythrocyte indices, 4-31
for hemoglobin synthesis, 4-25
for hemopoiesis, 4-18
for plasma volume, 4-118
tritiated thymidine, 4-13
Techniques, clinical
for visual acuity, 8-107
rotating (Barany) chair, 8-55
Techniques, endocrine
for neurohumoral controls, 7 13
hormone assays, general, 7 3
^{131}I uptake in thyroid gland, 7 39
identifying hormones, 7 1
tests of thyroid function, 7 33
Techniques, endocrine assay
ACTH, 7-61
calcitonin, 7 48
catecholamine, 7 72
corticosteroid, 7-53
insulin, 7-76
releasing factors, 7-15
renin, 7-62
vasopressin, 7-25
Techniques, general
for acid-base balance, 4-111
for fluid volumes in body, 4-113
for intestinal absorption, 2-83
for study of membranes, 1-3
for ultrastructure of muscle, 1-74
hypothalamic heating, 9-134
problem-solving, 9 155
radioautography, 4-13
thermopile for muscle, 1-85
Techniques, genetics, chromosome mapping, 1-162
Techniques, neurological
electrical recording, 9-9
electrical stimulation, 9-7
electromyography, 1-104
for action potential, 9-87
for clinical reflexes, 9-87
for spinal reflexes, 9-70
for synaptic potentials, 1-53
implanted electrodes, 9-6
microelectrodes, 9-10
stereotaxic surgery, 9-5
voltage clamp, 1-47
Techniques, physical, for light, 8-65
Techniques, psychophysical, 8-2 ff
for hearing, 8-31 ff
for hearing impairment, 8 35
for olfaction, 8-139
for taste, 8 128
for temperature sense, 8-152
for vestibular function, 8-49
retinal sensitivity, 8-85
Tectal autonomic outflow, 9 57
Tectorial membrane of cochlea, 8-15
Tectospinal tract, 9 29
Teeth, in chewing, 2 95
Tegmentum of brainstem, 9 30
Telencephalon, 9-20
Temperature
and central signal generation, 9 16
effect on food intake, 2 5
effect on gases, 6-2
effect on respiration, 6 52
effect on taste, 8-129
Temperature regulation, 9 131
and properties of blood, 4 1
endocrine mechanisms, 9 136
Temperature sense, 8-152 ff
hypotheses concerning, 8-154
Temperature vs metabolic rate, 9-133
Temporal lobe, 9 110
Temporal modulation transfer function, 8 111
Temporal summation, 9 76
Temporary threshold shift, 8-21
Tenderness, visceral, 2-12
Tendon organs (Golgi), 9-84
Tendons, functions of, 1 107
Tenotomy and differentiation of muscle, 1 112
Tension of resting muscle, 1 82
Terminal cisternae in muscle, 1 70
Termination of peptide chain, 1-181
Termination of polypeptide chains, 1-176
Tertiary structure of protein, 1-125
Testis, control of, 7-95 ff
Testis tubules, 7-99
Testosterone (see also Androgen)
and sex differentiation, 7-93
binding to globulin, 7-98
binding to nuclear receptor, 7-99
secreted by testis, 7-95
transport and metabolism, 7-97
Testosterone secretion in female, 7-106
Tetanus
isometric, 1 82
isotonic, 1-80
Tetralogy of Fallot, causes hypoxia, 6-77
Thalamic animal, 9-94
Thalamic lesions, 9-166
Thalamic syndrome, 9-167
Thalamocortical projections, 9-108
Thalamus
anatomy of, 9-39
and arousal system, 9-40
association functions, 9-40
nonspecific nuclei, projections, 9-108
sensory connections, 9-40
specific systems, 9-108
Thalassemia, 4-36
Therapy, oxygen, 6-81 ff
Thermal conductivity of blood, 4 1
Thermal detection in skin, 8 152
Thermal sensation in gut, 2-10
Thermal stimulation of vestibule, 8-49
Thermopile for muscle, 1-85
Thermostat, hypothalamic, 9-135
Thermostromuhr, measurement of blood flow, 3-156
Thiocyanate vs I in thyroid, 7-30
Thiourea, action in thyroid, 7 32
Thirst, 2-6 ff
Thiry-Vella loop, 2-83
Thoracic duct, lymphocytes in, 4-95
Thoracolumbar system, 9 53
Thorax in respiration, 6 28 ff
Three-channel hypothesis, 1-48
Threshold, neural, 1 37
for hearing, 8-31
for olfaction, 8-139
for taste, 8-128
for touch, 8-151
in vestibular system, 8-50
Threshold, renal, 5-12
Threshold shift, temporary, 8-21
Threshold tilt angle, 8-52
Thrombin, 4-75
Thrombin as coagulant, 4 82
Thrombocytes, 4-73 ff
Thrombocytopoiesis, 4 74
Thromboplastin, 4 75
Thromboplastin generation test, 4 79
Thrombosis
in sickle cell anemia, 4 37
pathogenesis of, 4-80
Thrombosthenin, 4-73
Thymectomy and immune response, 4-92
Thymic alymphoplasia, 4-102
Thymidine, tritiated, 4-13
Thymine, 1 142
Thymine dimer formation, 1-168
Thymus
in immunity, 4-92 ff
migration of cells, 4-96
source of lymphocytes, 4-95
Thyroglobulin, 7-32
transportation through thyroid cells, 7-33
Thyroid binding, abnormal, 7 37
Thyroid C cells, 7 46
Thyroid colloid, 7-33
Thyroid deiodinase, 7-33
Thyroid disorders, 7-29
Thyroid gland, 7-29 ff
and regulation of body temperature, 9-136
control of function, 7-29 ff
Thyroid hormones
blocking of synthesis, 7-32
distribution and metabolism, 7-33
increase in coronary blood flow, 3-138
mechanisms of action, 7-38
stored as thyroglobulin, 7-32
synthesis, storage and release, 7-30
vs dinitrophenol, 7-39
Thyroid pituitary-axis, 7-36
Thyroid stimulating hormone (see TSH), 7-35
Thyrotrophs of anterior pituitary gland, 7-9
Thyrotropin (see TSH), 7-35
Thyroxin, 7-29
Tic douloreux, 9-35
Tickling, 8-157
Tidal volume of respiration, 6-31
changes in emphysema, 6-74
changes in pneumonia, 6-71
Tight junctions, 1-14
Tilt angle, threshold, 8-53
Time factors
vestibular, 8-53
visual, 8-110
Tinnitus with deafness, 8-35
Titratable acid of urine, 5-30

Titration curve
of amino acid, 1-119
of buffer systems, 4-104
renal absorption of glucose, 5-11
T_m for glucose, 5-10
Tolbutamide and insulin release, 7-80
Tolerance, immunological, 4-92
Tone of muscle, 9-89
Tone vs spasticity of muscle, 9-166
Tongue in taste sensation, 8-122 ff
Tonic innervation from cerebral cortex, 9-116
Tonic postural reflexes, 9 95
in decerebrate state, 9-91
Tonometers for blood gases, 6-8
Tonotopy of auditory system, 8-30
Total peripheral resistance, 3-23
Touch sensation, 8-145 ff
in gut, 2-10
Toxic nodule in thyroid gland, 7 41
Tracheobronchitis from O_2, 6-82
Tracts, neural (see also name of tract or pathway)
myelination of, 1-31
Tracts, spinal and brainstem, 9-24
Tractus solitarius
and facial nerve, 9-35
and nerve IX, 9-37
and nerve X, 9-38
receives taste fibers, 8-128
Transamination
in intestinal absorption, 2-90
of amino acids, 1-132
Transamination systems, 1-134
Transduction, neuroendocrine, 9-124
Transfer factor in immunity, 4-98
Transfer function
spatial modulation, 8 108
temporal modulation, 8-111
Transfer RNA (tRNA)
activation of amino acids, 1 173
and peptide chain initiation, 1 179
in control of protein synthesis, 7-149
Transferase, glycogen synthesis, 7-117
Transferrin, 4-39
Transferrin in plasma, 4-6
Transformation of bacteria, 1-163
Transfusion (see also Blood transfusion), 4-48
cardiac circulation effects of, 3-137
Transitions cause mutations, 3-183
Translational controls for genes, 1-191
Translocation, peptide synthesis, 1 181
Transmembrane potential (see also Membrane potential, Potentials), 1-34 ff
Transmission, chemical, 1-58 ff
CNS transmitters, 9-14
distribution of CNS transmitters, 9-83
in autonomic system, 9-59
in basal ganglia, 9-100
Transmission, electrical, 1-59
Transmission, electrical vs. chemical, 9 82
Transmission of sound, 8-10
Transmission, synaptic, 1-52 ff
Transplantation of cells, 4-19
Transport
by renal tubules, 5-8 ff
of amino acids, 1-139
of glucose, facilitated, 7-114
of plasma lipid, 4-4
of sugars into muscle, 7-115
Transport, active, 1-21
Transport, carrier-mediated, 1-20
Transport, uphill vs downhill, 1 22
Transverse gyrus of Heschl, 9 110
Transverse tubules
in activation of muscle fiber, 1-90
of twitch vs slow fibers, 1-111
ultrastructure of, 1 70
Transversions cause mutations, 1-183
Trapping of I^+ by thyroid, 7-30
Tremor
as rhythmic mechanism, 9-102
in cerebellar dysfunction, 9 97
of Parkinson's disease, 9-99
TRF (see alse TSH releasing factor), 7-16
Triad of sarcoplasmic reticulum, 1-70
Tricarboxylic acid cycle
and energy release in muscle, 1-101
in amino acid metabolism, 1-132
in mitochondria, 1-11
in nerve, 1-51
Triceps jerk, 9-87
Trichromacy, visual, 8-113
Trigeminal nerve (V), 9-33 ff
and oral sensation, 8-126
Trigeminal neuralgia, 9-35
Triglyceride absorption, 2-91
Triglyceride homeostasis, 7-126
Triglyceride lipase, 7-129
Triglycerides, 7-123
Triiodothyronine, 7-29
Triplet-degenerate code, 1-178
Tritiated thymidine
study of granulocytopoiesis, 4 66
study of hemopoiesis, 4-13
study of lymphocytopoiesis, 4-93
Trochlear nerve (IV), 9-32
Tropomyosin
concentration in myocardium, 3-32
interaction with actin and myosin, 3-38
Tropomyosin and troponin, 1-66
Troponin
concentration in myocardium, 3-32
interaction with actin and myosin, 3-38
Troponin and Ca ions, 1-66
Trypsin, pancreatic, 2-51
Trypsinogen, pancreatic, 2-51
Tryptophan and protein synthesis, 7-151
Tryptophan operon, 1-189
TSH releasing factor (TRF), 7-16
chemical structure, 7-18
synthesis and release, 7 36
TSH (thyrotropin), 7-35
control of secretion, 7-15
secreted by placenta, 7-109
Tuber cinereum, 9-41
Tubules, renal, 5-1 ff
ammonia excretion, 5-31
bicarbonate transport, 5-27 ff
chloride transport, 5-23 ff
excretion of drugs, 5-37
H^+ secretion, 5-29
measurement of function, 5-40
non-ionic diffusion, 5-37
obligatory vs facultative water reabsorption, 5-19
potassium transport, 5-26
reabsorption in, 5-8 ff
secretion by, 5-13 ff
sodium transport, 5-23 ff
Tumors of adrenal medulla, 7-75
Tuning of neural circuit, 9-102
Turnover of hemopoietic cells, 4-19
Turnover of protein, 7-147
Turnover of red cells, 4-24
Turnover rate of platelets, 4-74
Turnover time in erythropoiesis, 4-23
Twitch muscle fibers, 1-110
Two-neuron arc, 9-1
inhibition of, 9-80
Two-point threshold, 8-151
Tympanic membrane, 8 12
Tympanic temperature, 9-132 ff
Tyrosine
iodinated derivatives, 7 30
precursor for catecholamine biosynthesis, 7-71
precursor of thyroid hormones, 7-29
Tyrosine transaminase activity, 7 149

Ubiquinone in muscle, 1-102
UDPG pathway in muscle, 1-98
Ulcerative colitis, 2-117
Ultrafiltration, renal, 5-4
Ultrasound, measurement of blood pressure, 3-152
Unidirectional vs net flux, 1-25
Unit cell membrane, 1-3
Units of CHO metabolism, 7-112 ff
Units of light, 8 66
Unmyelinated fibers, 1-31
Uphill vs downhill transport, 1-22
Uracil, 1-142
Urea
conservation in ruminants, 5-34
excretion during diuresis, 5 35
in uremia, 5-46
induces denaturation, 1-128
renal excretion, 5-34 ff
Urea clearance, 5-39
Urea cycle, 1-153
Urease in gastric juice, 2-38
Uremia, 5-45, 46
Ureogenesis and glucagon, 7 85
Ureters, 5-48 ff
Urethra, 5-49
in micturition, 5-55
Uric acid
from degradation of purine, 1-151
transport in kidney, 5-13
Uridine-5'-monophosphate, 1-143
Uridine ribonucleotide, 1-144
Urinary bladder, 5-47
physiological capacity, 5-53
Urinary corticosteroids, 7-55
Urination, 5-54, 55
Urine flow
and osmolarity, 5-23
and solute excretion, 5-23
Urine formation, 5-1 ff
abnormalities of, 5-38 ff
acidification, 5-29 ff
control of specific activity, 5-21
hyper- vs hypotonicity, 5-20 ff
summary of processes, 5-3
urea excretion, 5-34 ff
Urine tests, 5-38 ff
Urobilin, 2-64
Urobilinogen, 2-64
Uropepsin, 2-37
Uroporphyrins, 4-42
Uterus, cycles in, 7-102
Utricles, 8-44
Uvula of cerebellum, 9-99

Vagus nerve, 9-37 ff
afferents affecting respiration, 6-48
and control of respiration, 6-43
and food intake, 2-4
and reflex vasopressin release, 7-26
autonomic outflow, 9-57
control gastric secretion, 2-35
control of hormone release, 2-16
in swallowing, 2-96
regulation of heart rate, 3-114
respiration after vagotomy, 6-45
taste fibers in, 8-126

to digestive tract, 2-8
to pancreas, 2-49
Vagusstuffe, 9-59
Valve, ileocecal, 2 116
van der Waal's bond in protein, 1-122
Van Slyke-Neill manometric apparatus, 6-9
Van Slyke volumetric apparatus, 6-9
Vascular tissue, anatomy, 3-164, 165
Vasoconstrictor fibers, 3-166 ff
Vasodilator fibers, 3 170, 171
Vasopressin (see also Antidiuretic hormone, ADH)
chemical structure of, 7-25 ff
control of release of, 7 25
neurosecretion of, 7-24
therapy of diabetes insipidus, 7-28
vasoconstrictor effects of, 3-172
Veins, pressure pulse of, 3-198
Velocity of light, 8-63
Velocity of sound, 8 10
Venoms cause hemolysis, 4-60
Venoms induce clotting, 4-82
Venous return, affects pulmonary ventilation, 6-49
Ventilation (see under Alveolar, Pulmonary ventilation)
Ventral nuclei of thalamus, 9-39
Ventral root potentials, 9-11
Ventral roots
spinal, 9-23
visceral fibers, 9 54
Ventral spinocerebellar tract, 9-25
Ventricles, heart
anatomy, 3-25
contraction, as determinant of venous blood pressure, 3-200
contraction force-velocity, 3-93 ff
conduction of impulse, 3-56
emptying, patterns of, 3-99 ff
equilibrium of left and right output, 3-124 ff
exercise effect on blood volume of, 3-119 ff
extrasystoles, 3-78
fibrillation, 3-83
pressure receptors, 3 185
pressure-volume during cardiac cycle, 3-105
transmembrane potential, 3-48
wall tension in contraction, 3-92
weight, 3-26
Vergence movement of eyes, 8-98
Vermis of cerebellum, 9-43
Vertigo, 8-59
Vesicles
embryonic brain, 9-21
membrane-delimited, 1-8
presynaptic, 1-52
Vestibular efferent system, 8-25
Vestibular function, 8-40 ff
Vestibular nerve, 8-49
cerebellar connections, 9-48
innervation of receptors, 8-46
Vestibular system
classes of responding cells, 8-58
control of eye movement, 8-98
electrical potentials in, 8-57
habituation in, 8-54
psychophysical study, 8-50
single units in brainstem, 8-58
temporal characteristics, 8-53
Vestibulocochlear nerve (VIII), 9-37
Vestibulospinal tract, 9 27
origin of, 8-49
Villi, intestinal, 2-114
Villikinin, 2-115
Virilization in adrenal hyperplasia, 7-67
Visceral motor areas, cerebral cortex, 9-116
Vis a tergo, pressure gradient in systemic circulation, 3-200
Visceral nerves, 9-54
Visceral nervous system (see also Autonomic nervous system), 9-53 ff
higher levels of control, 9-64
Visceral pain, 2-11 ff
Visceral reflexes to gut, 2-9
Visceral sensations from gut, 2-10 ff
Viscosity of blood, and suspension stability, 4-60
Vision, 8-63 ff
basic visual functions, 8-98 ff
energy requirements for, 8-85
photochemistry of, 8-81 ff
Visual acuity, 8-106
Visual cortex
activity of single units, 8-93
development of competence, 8-95
Visual deprivation, 8-95
Visual system
anatomy of, 8-70 ff
central pathways, 8 78
lateralization of function, 9-120
Visuosensory area, 9-111
Vital capacity, 6-31
at high altitude, 6-67
Vitamin B_{12}, 2-94
deficiency in pernicious anemia, 4-33
prevents pernicious anemia, 4-37
renal clearance of, 5-8
Vitamin D, 7-42 ff
question of renal action, 7-45
Vitamins, absorption of, 2-93
Vitreous humor, 8-72
Voice, communication via, 8-29
Voltage clamp technique, 1-47
Voluntary movement
and basal ganglia, 9-100
and cerebellum, 9-97
as rhythmic mechanism, 9-103
as signal generation, 9 16
Vomiting, 2-14
dehydration from, 4-116
mechanisms of, 2 110 ff
with intestinal obstruction, 2-116
von Békesy audiogram, 8-37
von Economo, Constantin, 9-104

Wakefulness, humoral control, 9-142
Waking state, 9-137 ff
Wallerian degeneration of nerve fibers, 1-32
Warm receptors, 8-153
Water (see also Water movement)
colligative properties, 5-17
distribution in body, 4-113
free water clearance, 5-22
heat loss in evaporation, 9-130
iceberg structure of, 1-123
structure of molecules, 1-16
Water balance, 4-114 ff
control of, 9-125
Water content of tissues, 4-114
Water depletion
and epileptic seizures, 9-172
causes, clinical features, 4 116
Water intake
and thirst sensation, 2-6
and water balance, 4-114
control mechanisms, 2-1
Water intoxication, 4-117
Water movement
in gastric secretion, 2-41
intestinal absorption, 2 84
obligatory vs facultative reabsorption, 5-19
osmosis, 1-27
permeability coefficient, 1-26
renal excretion, 5 17 ff
transmembrane, 1-17
Water output and balance, 4-114
Water retention in starvation, 4-116
Watson-Crick model of base pairs, 1-152
Watson-Crick model of DNA, 1-159
Wave on basilar membrane, 8 14
Wave theory of light, 8-63
Wavelength and color vision, 8-91
Wever-Bray phenomenon, 8-20
White blood cells (see name of cell, Granulocytes)
White muscle fibers, 1-109
White vs. red muscle, 9 90
biochemistry of, 7-112
Wilson's disease, 9-99
ceruloplasmin deficiency, 4-6
clinical description, 9-168
Wobble hypothesis for codons, 1 177
Wolf-Parkinson-White syndrome, 3-77
Wolffian duct, differentiation, 7-92
Work-induced growth, 7-147
Work of breathing, 6-36
Writing and speech, 9-148

X chromosome, 1-137
X-ray crystallography
of DNA fibers, 1-159
of nucleic acid crystals, 1-151
X-ray diffraction
of actin and myosin, 1-74
of membranes, 1-3
of peptide bond, 1-121
Xanthine oxidase, 1-152
Xanthinuria, 1-156
Xerostomia, 2-30

Y chromosome, 1-137

Z line, myocardial sarcomere, 3 30 ff
Zero gravity
adaptation to, 8-54
as test for vestibular function, 8-49
Zinc in insulin synthesis, 7-77
Zonula adherens, 1-14
Zwitterions, amino acids, 1-119
Zymogen granules, pancreatic, 2-54